THE AMERICAN JOURNAL OF PSYCHOLOGY

Vol. 82
1969

PUBLISHED QUARTERLY BY THE UNIVERSITY OF ILLINOIS PRESS
Urbana, Illinois 61801

Manufactured in the United States of America

TABLE OF CONTENTS

ARTICLES AND NOTES

ABOUD, J., JR., and GRANBERG, D., A Contextual Effect in Judgments of Numerousness ... 221
ADKINS, M., and DAVES, W. F., Stimulus Variation and Free Recall: A Confirmation ... 122
ANDREWS, O. E., LEVINTHAL, C. F., and FISHBEIN, H. D., The Organization of Hypothesis-Testing Behavior in Concept-Identification Tasks 523
ANSTIS, S. M., MAYHEW, J. W., and MORLEY, T., The Perception of Where a Face or Television 'Portrait' is Looking ... 474
ANSTIS, S. M., and WILSON, J. A., Visual Delay as a Function of Luminance ... 350
ANTONITIS, J. J., A New Photographic Illusion ... 389
BAER, P. E., and FUHRER, M. J., Cognitive Processes in Differential GSR Conditioning: Effects of a Masking Task ... 168
BAGLEY, S. K., CHUBB, N. C., HURST, P. M., and RADLOW, R., Effects of *D*-amphetamine on Acquisition, Persistence, and Recall ... 307
BECK, J., Lightness and Orientation ... 359
BENTON, R. G., MANCY, P., JR., ROBERTSON, B., and RAPPOPORT, D. A., A Two-Chamber Hurdle Apparatus for Using Odorants in Aquatic Animal-Training ... 133
BERGER, S. M., and ELLSBURY, S. W., The Effect of Expressive Verbal Reinforcements on Incidental Learning by Models and Observers ... 333
BEVAN, W., and COLLYER, S. C., The Role of Thematic Relevance of Context in Paired-Associates Learning ... 415
BITTERMAN, M. E., C. Lloyd Morgan and the Theory of Instrumental Learning ... 126
BOWMAN, R. W., JR., and MACKAY, D. G., On Producing the Meaning in Sentences ... 23
BREHMER, B., Cognitive Dependence on Additive and Configural Cue-Criterion Relations ... 490
BURNHAM, C. A., and ONO, H., Variables Altering Perception of the Rotating Trapezoidal Illusion ... 86
CAMPIONE, J. C., Intra- and Extradimensional Shifts in Retardates as a Function of Dimensional Preference ... 212
CHUBB, N. C., HURST, P. M., RADLOW, R., and BAGLEY, S. K., Effects of *D*-amphetamine on Acquisition, Persistence, and Recall ... 307
COLLYER, S. C., and BEVAN, W., The Role of Thematic Relevance of Context in Paired-Associates Learning ... 415
CRONNELL, B., SMITH, F., and LOTT, D., The Effect of Type Size and Case Alternation on World Identification ... 248
DAVES, W. F., and ADKINS, M., Stimulus Variation and Free Recall: A Confirmation ... 122
DAY, R. H., and WADE, N. J., The Reference for Visual Normalization ... 191
ELLSBURY, S. W., and BERGER, S. M., The Effect of Expressive Verbal Reinforcements on Incidental Learning by Models and Observers ... 333
ERIKSEN, C. W., and SCHURMAN, D. L., Summation and Interaction of Successive Masking Stimuli in Visual Perception ... 320
FILION, R. D. L., FOWLER, S. C., and NOTTERMAN, J. M., Psychophysical Evaluation of Feedback Phenomena as Related to Precision of Force Emission: Some Methodological Considerations ... 266
FISHBEIN, H. D., ANDREWS, O. E., and LEVINTHAL, C. F., The Organization of Hypothesis-Testing Behavior in Concept-Identification Tasks 523

FISHBEIN, M., LANDY, E., and HATCH, G., A Consideration of Two Assumptions Underlying Fiedler's Contingency Model for Prediction of Leadership Effectiveness 457
FOWLER, S. C., FILION, R. D. L., and NOTTERMAN, J. M., Psychophysical Evaluation of Feedback Phenomena as Related to Precision of Force Emission: Some Methodological Considerations 266
FREUND, J. S., and UNDERWOOD, B. J., Verbal-Discrimination Learning with Varying Numbers of Right and Wrong Terms 198
FUHRER, M. J., and BAER, P. E., Cognitive Processes in Differential GSR Conditioning: Effects of a Masking Task 168
GLASS, T. A., Effects of Three Situational Variables and Type of Reinforcement on Delay Preference 241
GLUCKSBURG, S., TARPY, R. M., and LYTLE, L. D., Motivation and Short-Term Retention: Evidence for Covert Rehearsal 111
GOGEL, W. C., Equidistance Effects in Visual Fields 342
GOGEL, W. C., The Absolute and Relative Size Cues to Distance 228
GRANBERG, D., and ABOUD, J., JR., A Contextual Effect in Judgments of Numerousness 221
GREENBERG, B., and LOVELACE, E. A., Cue Selection in Paired-Associate Learning: Length of Study Interval 254
HALPER, F., and ROCK, I., Form Perception without a Retinal Image 425
HAMILTON, D., and KENNEDY, R. A., Time to Locate Probe Items in Short Lists of Digits 272
HARCUM, E. R., Continuous Tachistoscopic Patterns Reproduced without Laterality Effects 504
HATCH, G., LANDY, E., and FISHBEIN, M., A Consideration of Two Assumptions Underlying Fiedler's Contingency Model for Prediction of Leadership Effectiveness 457
HURST, P. M., RADLOW, R., CHUBB, N. C., and BAGLEY, S. K., Effects of *D*-amphetamine on Acquisition, Persistence, and Recall 307
JOHNSON-LAIRD, P. N., and TAGART, J., How Implication is Understood 367
KAPLAN, M. F., and OLCZAK, P. V., Originality and Rate of Response in Association as a Function of Associative Gradient 157
KAPPAUF, W. E., An Empirical Sampling Study of the Dixon and Mood Statistics for the Up-and-Down Method of Sensitivity Testing 40
KAPPAUF, W. E., The Sixty-fifth Annual Meeting of the Society of Experimental Psychologists 547
KARKAU, V. T., SUMMERS, S. A., and SUMMERS, R. C., Judgments Based on Different Functional Relationships between Interacting Cues and a Criterion 203
KENNEDY, R. A., and HAMILTON, D., Time to Locate Probe Items in Short Lists of Digits 272
KOHFELD, D. L., Effects of Ready-Signal Intensity and Intensity of the Preceding Response Signal on Simple Reaction Time 104
LANDY, E., FISHBEIN, M., and HATCH, G., A Consideration of Two Assumptions Underlying Fiedler's Contingency Model for Prediction of Leadership Effectiveness 457
LANGER, J., ROSENBERG, B. G., and STEWART, J., Nonverbal Learning 181
LASKER, A., ROCK, I., and SIMON, J., Stimulus 'Generalization' as a Process of Recognition 1
LEVINE, M., MORGAN, R., and NEIBERG, A., A Methodological Study of Incidental and Intentional Learning 117
LEVINTHAL, C. F., ANDREWS, O. E., and FISHBEIN, H. D., The Organization of Hypothesis-Testing Behavior in Concept-Identification Tasks 523
LOTT, D., SMITH, F., and CRONNELL, B., The Effect of Type Size and Case Alternation on Word Identification 248
LOVELACE, E. A., and GREENBERG, B., Cue Selection in Paired-Associate Learning: Length of Study Interval 254
LOWRY, R., Galilean and Newtonian Influences on Psychological Thought 391
LYTLE, L. D., TARPY, R. M., and GLUCKSBURG, S., Motivation and Short-Term Retention: Evidence for Covert Rehearsal 111

MACKAY, D. G., and BOWMAN, R. W., JR., On Producing the Meaning in Sentences ... 23
MANDLER, G., Input Variables and Output Strategies in Free Recall of Categorized Lists ... 531
MAXCY, P., JR., BENTON, R. G., ROBERTSON, B., and RAPPOPORT, D. A., A Two-Chamber Hurdle Apparatus for Using Odorants in Aquatic Animal-Training ... 133
MAYHEW, J. W., ANSTIS, S. M., and MORLEY, T., The Perception of Where a Face or Television 'Portrait' is Looking ... 474
MORGAN, M. J., Estimates of Length in a Modified Müller-Lyer Figure .. 380
MORGAN, R., NEIBERG, A., and LEVINE, M., A Methodological Study of Incidental and Intentional Learning ... 117
MORLEY, T., MAYHEW, J. W., and ANSTIS, S. M., The Perception of Where a Face or Television 'Portrait' is Looking ... 474
MUNDL, W. J., A Food Dish for Testing Food and Stimulation Reward with Small Animals ... 406
NEIBERG, A., MORGAN, R., and LEVINE, M., A Methodological Study of Incidental and Intentional Learning ... 117
NEISSER, U., The Role of Rhythm in Active Verbal Memory: Serial Intrusions ... 540
NEWMAN, E. B., Newton, Physics, and the Psychology of the Nineteenth Century ... 400
NOTTERMAN, J. M., FILION, R. D. L., and FOWLER, S. C., Psychophysical Evaluation of Feedback Phenomena as Related to Precision of Force Emission: Some Methodological Considerations ... 266
OLCZAK, P. V., and KAPLAN, M. F., Originality and Rate of Response in Association as a Function of Associative Gradient ... 157
ONO, H., and BURNHAM, C. A., Variables Altering Perception of the Rotating Trapezoidal Illusion ... 86
RADLOW, R., HURST, P. M., CHUBB, N. C., and BAGLEY, S. K., Effects of *D*-amphetamine on Acquisition, Persistence, and Recall ... 307
RAPPOPORT, D. A., MAXCY, P., JR., BENTON, R. G., and ROBERTSON, B., A Two-Chamber Hurdle Apparatus for Using Odorants in Aquatic Animal-Training ... 133
ROBERTSON, B., BENTON, R. G., MAXCY, P., JR., and RAPPOPORT, D. A., A Two-Chamber Hurdle Apparatus for Using Odorants in Aquatic Animal-Training ... 133
ROBINSON, J. S., Familiar Patterns Are No Easier to See than Novel Ones 513
ROBY, T. B., Forecasting Event Cycles: A Laboratory Paradigm ... 374
ROCK, I., LASKER, A., and SIMON, J., Stimulus 'Generalization' as a Process of Recognition ... 1
ROCK, I., and HALPER, R., Form Perception without a Retinal Image ... 425
ROSENBERG, B. G., LANGER, J., and STEWART, J., Nonverbal Learning ... 181
ROYER, J. M., Associative Recall as a Function of Stimulus Recognition .. 96
RUMP, E. E., A Modified Replication of Cook's Invariance-Hypothesis Experiment ... 291
SCHIFF, W., and THAYER, S., Stimulus Factors in Observer Judgment of Social Interaction: Facial Expression and Motion Pattern ... 73
SCHURMAN, D. L., and ERIKSEN, C. W., Summation and Interaction of Successive Masking Stimuli in Visual Perception ... 320
SIMON, J., LASKER, A., and ROCK, I., Stimulus 'Generalization' as a Process of Recognition ... 1
SJÖBERG, L., The Dimensionality of Similarity Judgments ... 441
SMITH, F., LOTT, D., and CRONNELL, B., The Effect of Type Size and Case Alternation on Word Identification ... 248
STEGER, J. A., Visual Lightness Assimilation and Contrast as a Function of Differential Stimulation ... 56
STEWART, J., LANGER, J., and ROSENBERG, B. G., Nonverbal Learning ... 181
SUMMERS, D. A., Adaptation to Change in Multiple Probability Tasks ... 235

SUMMERS, R. C., SUMMERS, S. A., and KARKAU, V. T., Judgments Based on Different Functional Relationships between Interacting Cues and a Criterion 203
SUMMERS, S. A., SUMMERS, R. C., and KARKAU, V. T., Judgments Based on Different Functional Relationships between Intracting Cues and a Criterion 203
TAGART, J., and JOHNSON-LAIRD, P. N., How Implication is Understood ... 367
TARPY, R. M., GLUCKSBURG, S., and LYTLE, L. D., Motivation and Short-Term Retention: Evidence for Covert Rehearsal 111
TAYLOR, R. L., On Recognition as a Discrete Event 259
THAYER, S., and SCHIFF, W., Stimulus Factors in Observer Judgment of Social Interaction: Facial Expression and Motion Pattern 73
UNDERWOOD, B. J., and FREUND, J. S., Verbal-Discrimination Learning with Varying Numbers of Right and Wrong Terms 198
WADE, N. J., and DAY, R. H., The Reference for Visual Normalization ... 191
WADE, N. J., The Effect of Monocular and Binocular Observation on Visual Orientation during Head Tilt 384
WHITE, J. F., A Report on Myers' New Autokinetic Phenomenon 124
WILSON, J. A., and ANSTIS, S. M., Visual Delay as a Function of Luminance 350

BOOK REVIEWS

(The reviewer's name appears in parentheses after the title of the work.)

BARNES, L. B., DALTON, G. W., and ZALEZNIK, A., The Distribution of Authority in Formal Organizations (A. K. Korman) 554
BOULTER, L. R., ENDLER, N. S., and OSSER, H., Contemporary Issues in Developmental Psychology (M. Lewis) 552
CHERRY, C., On Human Communication: A Review, A Survey, and a Criticism, 2nd. ed. (M. Posner) 551
COHEN, J., Psychological Time in Health and Disease (J. A. Michon) 550
DALTON, G. W., BARNES, L. W., and ZALEZNIK, A., The Distribution of Authority in Formal Organizations (A. K. Korman) 554
ELLIS, N. R. (ed.), International Review of Research in Mental Retardation, Vol. II (J. Clausen) 139
ENDLER, N. S., BOULTER, L. R., and OSSER, H., Contemporary Issues in Developmental Psychology (M. Lewis) 552
FELLOWS, B. J., The Discrimination Process and Development (F. M. Rabinowitz and J. C. Campione) 142
HOOPER, F. H., and SIEGEL, I., Logical Thinking in Children (E. D. Neimark) 552
JACOBSON, E., Biology of Emotion (A. M. Mordkoff) 555
JUNG, J., Verbal Learning (J. Ceraso) 548
LITWIN, G. H., and STRINGER, R. A., JR., Motivation and Organizational Climate (O. W. Smith and P. C. Smith) 554
LOVEJOY, E., Attention in Discrimination Learning (D. H. Lawrence) 549
MACKAY, R. S., Bio-Medical Telemetry (W. H. Foy) 555
OSSER, H., BOULTER, L. R., and ENDLER, N. S., Contemporary Issues in Developmental Psychology (M. Lewis) 552
POWLEDGE, F., To Change a Child: A Report on the Institute for Developmental Studies (L. C. Kovar) 143
ROCK, I., The Nature of Perceptual Adaptation (Y. Kohler) 141
SIEGEL, I. E., and HOOPER, F. H., Logical Thinking in Children (E. D. Neimark) 552
SKINNER, B. F., The Technology of Teaching (C. N. Cofer) 549
STRINGER, R. A., JR., and LITWIN, G. H., Motivation and Organizational Climate (O. W. Smith and P. C. Smith) 554
ZALEZNIK, A., DALTON, G. W., and BARNES, L. W., The Distribution of Authority in Formal Organizations (A. K. Korman) 554

PSYCHOLINGUISTICS: SELECTED PAPERS

By Roger Brown, **Harvard University**

This collection of papers, spanning the 15 years from 1954 to 1969, reflects the changes in linguistic theory that have occurred in that time. "Part One: The Development of a First Language" includes the major papers on child speech, and "Part Two: Psycholinguistic Processes in Adult Life" includes psycholinguistics papers on miscellaneous topics. The selections range from early papers drawing upon structural linguistics to later papers reflecting the transformational linguistics of the mid-1960s.

February, 1970 — 250 pages — $8.95

AN INVITATION TO MODERN PSYCHOLOGY

By Howard F. Gallup, **Lafayette College**

Designed for introductory psychology courses, this text surveys basic concepts and provides ample bibliographic information for the further study it stimulates. Topics discussed range from personality, intelligence, and aptitudes to humor and prejudice. The lessons are relevant to current problems and are presented in a highly readable style.

1969 — 252 pages — paper, $2.95

READINGS IN HUMANISTIC PSYCHOLOGY

Edited by Anthony J. Sutich, **Editor, Journal of Transpersonal Psychology**, and Miles A Vich, **Editor, Journal of Humanistic Psychology**

Carefully balanced and offering an optimal amount of humanistically oriented psychological literature, this volume contains 24 articles presenting a cross section of developments and a wide range of views in this field. In addition to an original introduction and references following each article, this excellent text and source book includes a 140-item bibliography.

1969 — 455 pages — paper, $3.95

THE AMERICAN
JOURNAL OF PSYCHOLOGY

Founded in 1887 by G. STANLEY HALL

Vol. 82 MARCH 1969 No. 1

STIMULUS 'GENERALIZATION' AS A PROCESS OF RECOGNITION

By IRVIN ROCK and ALICE LASKER, Rutgers University, and JANET SIMON, Yeshiva University

An organism can respond to a stimulus if it identifies the stimulus as identical to or similar to one it has learned to respond to in the past, *i.e.* if it recognizes the stimulus. Psychologists of various persuasions would no doubt subscribe to this statement, but because the process of recognition has not itself been effectively dealt with, they do not seem to realize the implication of this view. The implication is that concepts such as 'association' or 'S-R connection' are not applicable to recognition. Indeed, experience with a single item can lead to a form of learning not encompassed by most current theories, because these theories have presupposed that learning *must* be associative. For although item-learning has been discussed in the literature, the assumption has been that it consists of integrating or associating the components of an item, *e.g.* of integrating the letters of a nonsense syllable. Even if it is true that nonsense syllables must be 'integrated' in this fashion, it is clearly not true that stimuli such as forms or any of the sensory qualities—*e.g.* color, tone, or odor—must be so 'integrated.' They need only be perceived.

In contrast to these theories, assume, rather, that an organism perceives a single stimulus—a form, a color, etc.—and that a memory of that stimulus is established. When, later, that stimulus or

* Received for publication July, 1968. A few of the experiments reported here were conducted when the first two authors were at the New School for Social Research and the remainder were conducted at Yeshiva University. The authors wish to thank John Ceraso, Morris Eagle, Hans Wallach, Eleanor Gibson, Sheldon Ebenholtz, and Solomon Asch for reading earlier drafts of this paper. This work was supported in part by a Research Scientist Award (K3-MH-31,361-01) from NIMH, USPHS, to the senior author.

one like it is presented again, the organism may recognize it as one previously encountered. To explain this process of recognition, it only seems necessary to say that the present stimulus makes contact with a memory trace of the relevant previous stimulus. The present stimulus may be similar, but not identical, to the previous one, and recognition may nevertheless occur;[1] the organism also may or may not be aware that the new stimulus is somewhat different from the previous one. Now, when A and B have been associated, recall of B presupposes that recognition of A has occurred. The reason for this statement is as follows: A and B were associated in the past. When A recurs (or a stimulus similar to A occurs), the perception of that stimulus is a new event. The present A is not associated with B; rather, the trace of the previously encountered A is associated with B. Therefore, the present A must first make contact with the trace of the previous A before we can reasonably expect the associated response B to occur. This was the point made many years ago by Höffding.[2] In the present paper, this process of the present A contacting the previous A is called 'recognition.' It produces a sense of familiarity, of having seen or heard A before, whether or not B is then recalled. B may not be recalled for a variety of reasons: the association between A and B is not sufficiently strong, B is momentarily unavailable, etc. But if recall of B is possible only when A is recognized, it then follows that any factor that leads to failure of recognition of A will, ipso facto, lead to failure of recall of B. If the stimulus A is sufficiently different from the previous A, the trace of A may not be aroused and recognition will not occur. In that event, recall of B will not occur. Therefore, failure of recall of B (or any evidence of attenuation of response B, *e.g.* increased latency or decreased amplitude) when the present A is changed may reflect the failure of or a change in the recognition of that stimulus. It is recognition that is most directly affected when the stimulus is changed, but since recall depends upon recognition, the recall or response measure employed to study 'generalization' mirrors what is happening with respect to recognition. The recognition is usually revealed by a response associated with the previous stimulus because we do not test recognition directly, at least in animals or infants. But as far as

[1] Köhler and others have argued that activation of the trace is on the basis of similarity to the present stimulus. See W. Köhler, *Dynamics in Psychology,* 1940. See also I. Rock, A neglected aspect of the problem of recall: The Höffding Function, in J. Scher (ed.), *Theories of Mind,* 1962.

[2] H. Höffding, *Outline of Psychology,* 1891.

stimulus 'generalization' is concerned, the associated response can be considered to be primarily an index of the fact that recognition has taken place.

In short, the present paper suggests that stimulus 'generalization' can be understood as the recognition of a stimulus similar but not identical to a previously encountered stimulus. In the realm of perception, we are accustomed to expect recognition or response equivalence when the stimulus is changed in certain respects; we speak, for instance, of transposition of a form or melody, by which we mean that the reaction will be the same because certain essential whole qualities are the same. Apparently then, the memory trace can be and is activated when the stimulus is only similar, but not necessarily identical, to the one originally seen. In fact, if recognition of stimulus A depends on the arousal of trace A on the basis of a similarity even when stimulus A is unchanged, then it would follow that identity is not required and that certain non-identical stimuli can activate the trace as well, *i.e.* that there is a tolerance of the trace-arousal mechanism such that a new stimulus is capable of activating the appropriate trace as long as it is sufficiently similar to the original stimulus. From this point of view, what we call stimulus 'generalization' is a predictable fact, not a problem to be explained. That is to say, the central problem is that of trace arousal on the basis of similarity, but recognition of non-identical and identical stimuli are not separate issues. It is the absence of 'generalization' that would be a problem. It may be observed at this point that since recognition may occur despite awareness of certain stimulus differences, recall cannot be expected to fail because the stimulus is different unless *S* does not identify that stimulus with the one used in training. 'Generalization' may be a function of *S*'s identifying the test stimulus with the training stimulus (as Lashley and Wade and as Prokasy and Hall imply),[3] but *S* need not necessarily be unaware of any difference. Where there is no awareness of difference, it would seem inappropriate to use a term such as 'generalization,' since there is then no reason to assume any process different from that which takes place when an identical stimulus is used; where there is awareness of difference and of similarity too, the term is perhaps less inappropriate; but in either case *S* is clearly not engaged in a process of universalizing

[3] K. Lashley and M. Wade, The Pavlovian theory of generalization, *Psychol. Rev.*, 53, 1946, 72–87; W. F. Prokasy and J. F. Hall, Primary stimulus generalization, *Psychol. Rev.*, 70, 1963, 310–322.

from particular instances. It is for this reason that the term 'generalization' is used in quotation marks throughout the paper.

It may be helpful to contrast the theory outlined above with the traditional association theory of stimulus 'generalization.' According to Clark Hull, for instance, an association or habit is established, during learning, not merely to the stimulus to which the response is conditioned, but to other similar stimuli along a continuum.[4] The habit has maximum strength for the reinforced stimulus and lesser strength for the similar stimuli; it decreases in strength the farther away the stimulus on the continuum from the reinforced stimulus and thus results in a gradient of generalization. On the other hand, the theory proposed in the present paper denies that multiple habits are established during learning. Only one association is formed. During the test, other stimuli can activate that association if they can activate the memory of the stimulus used in training, *i.e.* if recognition occurs. Thus according to this view, there is an act of perceptual recognition intervening in the chain of events at the time of the test, and this recognition is not just another response index of a habit, but a process distinct from habit.

The hypothetical internal events bearing on recognition and association are the aspects of stimulus 'generalization' with which the present paper is concerned. Obviously, there are many other facets of 'generalization' bearing on the organism's behavior, *e.g.* the roles of reinforcement and partial reinforcement, extinction, discrimination training, measures of response, and the like. Although the theory presented here may ultimately prove relevant to these aspects of 'generalization' behavior, this paper does not seek to deal with them. Rather, it seeks to show, first of all, that recall of B is contingent upon recognition of A when the latter is presented in a test (Experiment I). For although recognition and recall are generally treated as different methods of testing retention, and although recognition is known to be easier than recall, the two have not been considered interrelated in the contingent manner suggested here. Once this dependence of recall on prior recognition is established, the present paper tries to show that 'generalization' of an association pertains to the recognition of A, not to the association between A and B (Experiments II and III). Obviously, many difficulties with the theory must here remain essentially unanswered. Some of them are briefly considered in the Discussion.

[4] C. Hull, *Principles of Behavior*, 1943.

Experiment I

The most direct way of determining whether or not A must be recognized before the associated B can be recalled would be to ask S simultaneously to report on an item's familiarity and to recall the association to it. The prediction from the theory proposed would be: No recall without recognition, although recognition may occur without recall. Unfortunately, such a procedure might not lead to convincing results because S would not be likely to say, 'No, I do not recognize that,' while nevertheless giving an association to the item. Such behavior would seem contradictory to S, and he would doubtless try to avoid it. (The fact that it would seem contradictory attests to an intuitive realization of the contingency between recognition of A and the recall of an associated B. There is no a priori logical reason why one cannot recall B and not recognize A.) Hence one might argue that the predicted outcome cannot fail to occur. To get around this difficulty, recognition and recall were separated in time.[5]

Procedure. Immediately after S's exposure to a list of paired associates, the stimulus items alone were presented together with an equal number of new items of the same category, and S was asked to say 'yes' when he recognized a stimulus that had appeared in the learning period. After this recognition test, the stimuli from the original learning were presented again, and S was tested for recall of the associated items. The items to be learned were 48 pairs of first names and simple concrete nouns, each of one syllable, printed on 3 × 5 in. cards. The list was presented for one trial, at a rate of 5 sec. per card with a 1-sec. interval between cards. For the recognition test, the 48 names were shuffled together with 48 additional names; 3 sec. were allowed per card with a 1-sec. interval between cards. For the recall test, the 48 original names were presented again, and S was asked to give the correct noun associated with each; 8 sec. per card were allowed, with a 1-sec. interval between cards. A long list was used to ensure failures of recognition and to make it less likely that S would remember what he had said in the recognition test when subsequently tested for recall.

After five Ss had completed the experiment, six more Ss were run by the same procedure, except that in the recall test, they were asked to guess or to say the first thing that came to mind when they could not think of the correct noun. The first five Ss often made no response at all to the items presented for recall, perhaps because they were reluctant to respond to items to which they remembered saying 'no' in the recognition test. Forcing the last

[5] Either recall for an entire list can precede recognition for the entire list or vice versa. In a preliminary experiment we adopted the first procedure. We subsequently discarded it when we realized that, with recognition following recall, a response of 'yes' might signify recognition of an item from its presentation during recall (a second encounter) rather than from its presentation during initial learning.

six Ss to respond provided additional opportunity for negative cases, *i.e.* recall despite prior nonrecognition. The ample time for recall (8 sec.) and the use of names as stimulus items also favored negative cases: a name could, for instance, appear familiar at the time of recall, thus not inhibiting recall, even though it was not positively identified during the recognition test.

Results. Since 48 pairs were shown to each of the first five Ss, there were 240 items. Of these, 195 (81%) were recognized. Only 16 (6.5%) of the new items introduced on the recognition test were incorrectly identified as familiar. There were 100 (42%) correct responses on the test of recall. The crucial question, of course, was whether any of these correctly recalled items had not been previously recognized. Since 19% of *all* items presented for learning were not recognized, one might expect 19 of the 100 items correctly recalled not to have been recognized if there were no particular relationship between recall and recognition. In point of fact, there was only one such case. This was very much in accord with the prediction. In general, recognition was superior to recall, and there were many instances of recognition without recall. Of the 140 instances of no recall, 96 were previously recognized and 44 were not. Hence, virtually all (44 of 45) cases of no recognition were followed by no recall.

For the six Ss forced to respond in recall there were 288 total items. Of these, 202 (70%) were recognized. Only 17 (6%) of the new items introduced on the recognition test were incorrectly identified as familiar. There were 99 (34%) correct responses on the test of recall. Since 30% of *all* items presented for learning were not recognized, one might also expect 30 of the 99 correctly-recalled items not to have been recognized if there were no particular relationship between recall and recognition. There were actually five such cases. Again, recognition was superior to recall. Of the 189 instances of no recall, 108 were previously recognized and 81 were not. Hence, virtually all (81 of 86) cases of no recognition were followed by no recall.

The hypothesis that recall of B presupposes recognition of the associated A does not permit a single exception. Therefore if the few exceptions above could not be explained in any other way, the hypothesis would have to be considered disproven. However, the necessity of separating recognition and recall in time makes it difficult to be completely certain what the recognition status of an item was *at the time of recall*, when the item might have been recognized even though S had said 'no' during the recognition test. Less

time was allowed for recognition than recall; hence during the long 8-sec. exposure of an item during recall, recognition might have occasionally succeeded where previously it had failed. Furthermore, there was a conservative tendency operating during recognition: 30% of the original items were not recognized, whereas only 6% of the additional items were incorrectly recognized. In other words, for all items combined, *Ss* said 'yes' 38% of the time and 'no' 62% of the time, which suggests that when there was some doubt, *S* said 'no.' Finally, it is possible that some cases of correct recall were the result of guessing. For these reasons, the few exceptions can be tolerated without being considered to contradict the hypothesis that recall is contingent upon recognition. Hopefully, a technique may eventually be found that will rule out these alternative explanations of negative instances.[6]

Experiment II

If recognition of A is a necessary precondition for recall of B, then any 'generalization' which occurs in the recognition of stimuli similar to A would make it possible for the associated B to be recalled. (This assumes that once the trace of A is aroused, B has as good a chance of coming to mind when stimulus A is only similar to trace A as when it is identical. Of course, recall of B may fail to occur even if A is unchanged, so that complete 'generalization' means recall for similar stimuli is as good as for the identical stimulus, not that recall is perfect). Any failure of 'generalization' in the recognition of such similar stimuli would lead to failure of recall of B. In other words, if it can be shown that 'generalization' or the lack of it is a matter of recognition or failure of recognition (involving A alone), then it would be gratuitous to assume that 'generalization' has anything to do with the association between A and B. On this reasoning were based Experiments II and III. Briefly,

[6] After this paper was completed, studies reported testing the hypothesis that associative activation was contingent upon recognition of the stimulus: see E. Martin, Stimulus recognition in aural paired-associate learning, *J. verb. Learn. verb. Behav.*, 6, 1967, 272–276; E. Martin, Relation between stimulus recognition and paired-associated learning, *J. exp. Psychol.*, 74, 1967, 500–505; H. A. Bernbach, Stimulus learning and recognition in paired-associate learning, *J. exp. Psychol.*, 75, 1967, 513–519. In these experiments recognition and recall for a given item were tested at the same time, thus not avoiding the problem described above. In fact, in the studies by Martin, *S* was not told to try to recall if he considered a stimulus "new" or even to "guess" but only to choose one of three numbers (in one experiment) or (in the other experiment) to say any of eight digits that came to mind, when these eight were the associated terms for all items. Apart from this possible flaw, however, these studies indicate that recall of B depends upon recognition of the associated A.

all *Ss* were required to learn paired associates and were then tested for recall with the stimulus member of one critical pair either changed or left the same. The proportion of correct recall computed for groups given different critical stimuli then provided the kind of data often obtained in 'generalization' studies. *Ss* were also tested for their recognition of the critical stimulus. The prediction was that for certain changed stimuli, recognition and, therefore, recall would be unimpaired or, to put it another way, 'generalization' would be complete. For other changed stimuli, recognition would not occur, which would necessarily lead to a decrement in recall and thus yield a gradient when the data for all *Ss* were combined. The experiments were thus designed to show that the absence or presence of a decrement in recall of the item associated with the critical stimulus was completely explicable in terms of the presence or absence of recognition of that stimulus.

Materials. In two preliminary studies, the critical stimulus was a circle which was varied in size, or an ellipse which was varied in the ratio of its axes, for different groups during the test. Then, because it is difficult to find a figure which will yield a gradient when it is changed along some specifiable continuum, it was decided to make use of the well-known fact that most shapes look different when seen in different orientations. It is obviously possible to specify degree of change in terms of degree of change in orientation, but it is questionable if this is the relevant psychological variable. There is no physical change of the stimulus involved when it is differently oriented. It is true the proximal stimulus, the retinal image, is increasingly changed from its original position the more the figure is tilted, but it has been demonstrated that that is not the factor which produces the phenomenal change. The orientation of the retinal image can be held constant—by techniques such as tilting the observer the same amount as the figure—and still the figure will look different in orientations which differ with respect to the up and down of space.[7] Furthermore, as was demonstrated for the figure employed here, phenomenal change is not necessarily a function of degree of orientation change. For these reasons, orientation was a dimension ideally suited to the present purposes. Based on facts in the realm of perception, the stimulus could be modified to alter its phenomenal appearance, thereby altering its recognizability and so influencing the probability of recall. Fig. 1 shows the critical stimulus so devised (*A*) and its variations (*B*, *C*, and *D*). *B* represents a 90° tilt clockwise; *C* a 114° tilt clockwise; *D* a 34° tilt counterclockwise. Although *B* is tilted 90°, there are reasons for believing it is perceived as more similar to *A* than *D*, which is only tilted 34°. *B* is symmetrical about a horizontal

[7] I. Rock, The orientation of forms on the retina and in the environment, this JOURNAL, 69, 1956, 513–528; I. Rock and W. Heimer, The effect of retinal and phenomenal orientation on the perception of form, this JOURNAL, 70, 1957, 493–511; I. Rock and R. Leaman, An experimental analysis of visual symmetry, *Acta Psychol.*, 21, 1963, 171–183.

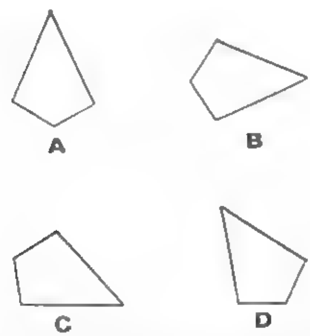


FIG. 1. CRITICAL FIGURES: EXPERIMENT II

axis: A is symmetrical about a vertical axis. These are the two important directions for axes of symmetry.[8] Also, B rests on a point, as does A. The other figures, C and D, are so oriented that a side is horizontal, forming a base. Viewed naively (no longer an easy achievement for the reader), the symmetry of the base with the side extending upward from the base on the right side is completely absent from experience. Hence C and D appear as irregular, asymmetrical quadrilaterals, with one side resting on the ground. Possibly C will be recognized somewhat less frequently than D because it rests on a longer base than does D, not because it is tilted 114°. It was expected, therefore, that B would be recognized a high percentage of the time, although not quite as often as A itself; that C and D would be recognized much less often than A or B; that C might be recognized less often than D.

Procedure. *Ss* were presented with six slides (in the preliminary studies) or six cards to learn, each containing a figure and a nonsense syllable. One of the six was the critical figure (circle, ellipse, or A). The other five, noncritical, figures were rectilinear geometrical shapes. The nonsense syllables paired with the six figures were randomly varied from S to S. The series was presented twice in random order. Cards were shown for 5 sec. each. *Ss* were instructed to learn the association between each figure and its syllable. Following the second trial all *Ss* were given a mimeographed article to read for 5 min. The *Ss* in the main experiment were college students randomly assigned to one of four groups of 25 *Ss* each.

In the recall test, one group of *Ss* was presented with the critical figure (standard circle or ellipse in the preliminary studies; A in the main experiment), whereas three other groups were given critical figures that were variations of circle size or ellipse axes (preliminary studies) or B, C, and D respectively. The remaining five test figures for all groups were the noncritical figures seen in training. The critical figure was always presented first in the test, and the order of the remaining five, noncritical, figures was the same for all groups. (Any other position of the critical figure in the test would not be desirable because after exposure of the other training stimuli, a set would exist to the effect that the training stimuli were being shown without modifi-

[8] E. Goldmeier, Ueber Aehnlichkeit bei Gesehen Figuren, *Psychol. Forsch.*, 21, 1937, 146–208.

cation. Also, *S* might even begin to anticipate the critical stimulus by a process of elimination.) As each stimulus was exposed, *S* had 8 sec. to write down the associated nonsense syllable. This part of the procedure is similar to that first used by Yum (1931), and later by Gibson (1941), to study generalization in humans by means of paired-associates learning.[9] One very important difference, however, is that Yum and Gibson varied the stimuli for many or all of the test items, whereas the present investigators varied only one, the very first presented in the test. We suspected that if many test items were varied, *Ss* would quickly realize that the stimuli they had seen in training were being presented in some modified form—a state of mind hardly congruent with the goal of tapping the spontaneous reaction to changed stimuli. In fact, we had performed an experiment like Gibson's and found that *Ss* did develop such a realization and that it led to various response strategies. Thus we sought to avoid any definite expectation, either that unchanged stimuli were being shown or that changed stimuli were being shown, by using only one critical stimulus and presenting it first in the test. There were other differences in procedure from Yum's or Gibson's, *e.g.* number of pairs, number of training trials and interval between training and test.

After the recall test for each figure, *S* was tested for his recognition of the stimulus. No mention was made of the recognition task until recall for the first item was completed. Then *S* was asked to say 'same' if the figure looked exactly like one of those used in the training series, 'similar' if it looked somewhat changed to him, and 'different' if it did not look at all like one of the figures he had been asked to learn. He was given 5 sec. to make this response. The wording of this recognition test was such as to ensure *S*'s understanding that his response was to indicate his recognition or lack of it at the time he had seen the critical figure, moments earlier, during the test for recall. To be sure that *S* was properly identifying the critical figure when he said 'same' or 'similar,' he was asked to make a brief sketch of the corresponding figure he had been shown in training. It will be noted that in this design, recall and recognition for each item were tested together, unseparated by testing recall for the entire list and only then testing recognition, or vice versa as in Experiment I.

Results. In the preliminary studies, recall was about as good for the groups tested with figures of different size or shape from the training figure as for the groups tested with the training figure itself, *i.e.* 'generalization' was complete. However, recognition revealed no gradient either. Recognition of the altered test figures was virtually total; from what is known about perception, one would expect recognition for varying sizes and similar-shaped ellipses to be rather good. Thus, the groups did not differ in recognition, nor did they differ in recall, with the exception, in the ellipse study, of one group tested with a circle and another tested with an

[9] K. S. Yum, An experimental test of the law of assimilation, *J. exp. Psychol.*, 14, 1931, 68–82; E. J. Gibson, Retroactive inhibition as a function of degree of generalization between tasks, *J. exp. Psychol.*, 28, 1941, 93–115.

ellipse, the axes of which were in a ratio reversed with respect to the training ellipse. For these groups, recall dropped and so did recognition. The preliminary studies are not described in detail because the almost complete recognition of the altered test figures did not afford a sufficiently decisive test of the hypothesis. That is, the results clearly showed that 'generalization' in recall is correlated with successful recognition, but they did not adequately show that failure of 'generalization' is due to failure of recognition.

For the main experiment, Fig. 2 gives the proportion of *Ss* in each of the four groups that recalled the association for the critical pair. It can be seen that the two groups given figures *A* and *B* did not differ appreciably from one another, whereas the group given figure *D* recalled less than the control (or *A*) group. This difference was significant at the .05 level using a one-tailed test of χ^2. Recall for the group given figure *C* was probably spuriously high because this one group was superior in ability to the other three.

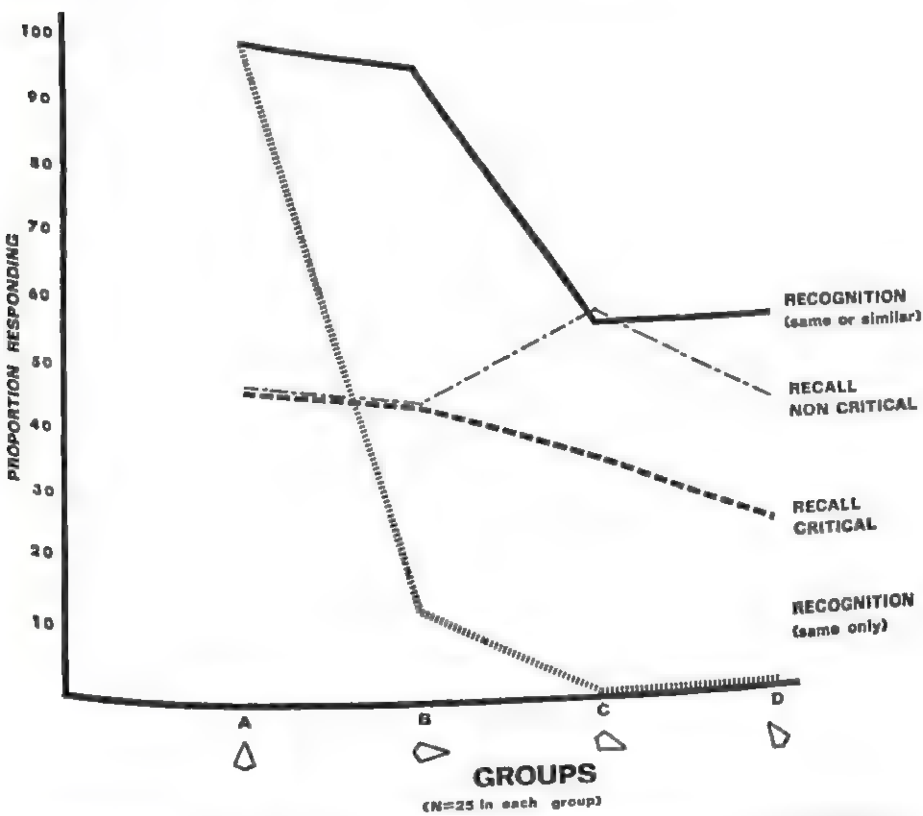


Fig. 2. Recall (Critical and Noncritical) and Recognition (Same or Similar and Same Only): Experiment II

Fig. 2 also shows that the group given C achieved a higher score on recall of noncritical items than any of the other groups. When this difference in ability was taken into account, recall for the critical item approached that of the group given D. Without taking this into account, the group given C was not significantly poorer than the control group, but if the data for the groups given A and B are combined and compared with the data for the groups given C and D, the two distributions were significantly different at the .05 level using a one-tailed test of χ^2. In other words, the critical item for the groups given C and D was less frequently recalled, *i.e.* a gradient of recall was obtained. (The overall level of recall of the noncritical figures was not different from that of the critical figure, at least for the control group and the group given B.)

The recognition curve for the four groups makes the recall data readily understandable. Recognition (responses of 'same' or 'similar') was essentially total for the groups shown A and B, whereas it dropped sharply for the other two groups, both of which differed significantly from the control group (which was shown A). This drop in recognition can account for the drop in recall for the groups that saw C and D. Of the cases of *successful* recognition for all groups, the proportions of these which led to recall showed no gradient: they were 48%, 46%, 57%, and 43% for the groups seeing A, B, C, and D respectively. Thus it is clear that the groups were essentially alike in their recall if one assumes that the pool of possible recallable items was composed of only those items which were recognized, and if one takes into account the superior ability of the group shown C. (The overall superiority in all groups of recognition to recall was, of course, to be expected.) In other words, there was no gradient in recall when recognition was successful. Hence it is fair to conclude that the gradient of recall when it was obtained was produced by the occasional nonrecognition of C and D, as this nonrecognition was expected on the basis of knowledge about phenomenal changes in perception. It is of interest that figures C and D led to the same decrement in recognition despite the fact that D represents a change of 34° and C a change of 114°. Figure B, which represents a change of 90°, did not show a decrement at all.

It is possible to go further than merely predicting group performance in recall on the basis of group data on recognition. In accordance with the contention that recognition of the test-stimulus A is a necessary condition for recall of B, there should also be no

recall when S says 'different' in the recognition test. In point of fact, except for one instance, this was the case, *i.e. where there was no recognition, there was no recall.* Of particular interest is the fact that these failures of recall were more often errors of no response than incorrect guesses, whereas the failures of recall where there was recognition (*i.e.* S said 'same' or 'similar' in the recognition test) were generally errors of incorrect response. The failure of recall in cases of no recognition and the tendency of these cases to be no-response errors were also found in the preliminary study with ellipses.[10] Recognition of the noncritical figures in the main experiment was uniformly high, there being only 13 failures out of the total of 500 test presentations to all Ss combined. Interestingly enough, none of these failures was preceded by correct recall, and 12 of the 13 failures were no-response errors.

It will be recalled that S was asked to draw the original figure from memory whenever he responded 'same' or 'similar' in the critical recognition test; in every case he was able to draw the figure with reasonable accuracy. Hence it cannot be claimed that correct recognition was merely an attempt to be consistent with a prior recall response. Nor, for two reasons, can it be argued that no recognition after no recall was also essentially an attempt at consistency. (1) There were many cases of no recall and then recognition, *i.e.* recognition is an easier task; there was no reason why S would feel inconsistent in recognizing where he failed to recall.[11] (2) All

[10] In Experiment I the recall errors can be analyzed in terms of no response versus incorrect response. For the first five Ss, no stress was laid on guessing, and most errors were of no response. However, when the name was not recognized, 93% of the errors were of no response, whereas 80% were of this kind when the name was recognized. For the second six Ss, no-response errors were ruled out by instruction. It is interesting to note, however, that when names were not recognized, the error in recall was more frequently not a noun from the list (84%), whereas when items were recognized, the error in recall was less frequently of this type (55%). It is as if Ss felt it inappropriate to guess an association from the list if the test item was not itself recognized as from the list.

[11] A somewhat different argument is that the correlation between recognition and recall merely reflects the fact that recognition depends upon recall, *i.e.* that stimulus recognition is a consequence of recall. It might be held that during a recognition test S covertly tries out responses and when he thinks he makes a correct response, he then says he recognizes the stimulus. The most convincing refutation of this argument is the fact that in all experiments where *both* recognition and recall have been tested—as in the experiments reported here—recognition often succeeds where recall fails. The fact is, of course, perfectly compatible with the theory suggested here because the activation of trace A on the basis of similarity of the stimulus need not lead to the activation of the associated trace B. Furthermore, Martin (*op. cit.*, *J. exp. Psychol.*, 71, 1967, 500–505) has provided evidence against the hypothesis that recognition depends upon recall.

but one case of no recognition occurred in the groups given figures *C* and *D*, *i.e.* in the groups for which one might predict a decline in recognition on the basis of perceptual change.

Does 'generalization' mean unawareness of any difference between training and test stimulus as Lashley suggested? No, because the plot of 'same' responses reveals a sharp gradient. In fact, only three of the 75 *Ss* in the three groups given a modified critical figure said 'same.' Yet many of these *Ss* did recognize the similarity of the figure and did recall the association. This sharp gradient of 'same' responses also appeared in the studies with circles and ellipses, where 'generalization' was virtually complete. Hence we see that 'generalization' is not a function of failure to discriminate.

Experiment III

Materials. Experiment III was performed in an attempt to produce phenomenal changes (in a figure) more directly correlated with physical changes. This was accomplished by rotating a cardboard cutout of a figure about a horizontal axis and tracing the projection of the shadow of the figure on a frontal plane. The figure itself was the mirror image of one of Gibson's standard figures shown in a different orientation. Fig. 3 shows *A* (the standard), which had a 54° acute angle on the left side, and *B* and *C*, for which the comparable angles were 32° and 16° respectively. The noncritical figures were polygons selected to be of comparable complexity.

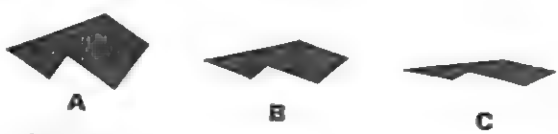


Fig. 3. Critical Figures: Experiment III

Procedure. The procedure was similar to that of Experiment II except for these changes. (1) Words, rather than nonsense syllables, were paired with the figures to make the learning task easier. These four-letter, one-syllable words were equated for frequency of appearance.[12] (2) There were two conditions: for half of the *Ss*, recall preceded recognition; for the other *Ss*, recognition preceded recall, to obtain spontaneous reports of each *S*'s perception of the critical figure, *i.e.* reports uncontaminated by intervening recall. There were three groups in each experimental condition: in the tests of recall and recognition, one group was shown the critical figure (*A*) as it appeared in training; another group was shown *B*; and the third group was shown *C*. (3) The training instructions were altered slightly so that the *Ss* getting the recognition test first would not be surprised by their task. All *Ss* were instructed: "Now I am going to show you some figures alone. I want you to

[12] E. L. Thorndike and I. Lorge, *The Teacher's Word Book of 30,000 Words*, 1944.

look at them, and then I will ask you some questions about them." The figure was then presented. *Ss* in the recognition-first condition were asked: "As you looked at that figure a moment ago, did you regard it as one of the figures you saw when you looked at the cards the first time? That is would you say it was identical to, similar to, or different from one of the figures you saw before?" If *S* said 'same' or 'similar,' he was asked to draw the figure. He was then asked: "Do you remember the word that went with it?" The instructions for the recall-first condition simply reversed the order of the questions. (4) Instead of having *Ss* read an article between training and the tests, they were given a digit-symbol task to perform.

During the test, 8 sec. were allowed for recall. Of course, for the recognition-first condition, the time between the initial presentation of the stimulus and recall varied, depending on whether *S* recognized the figure and on how long it took him to make the drawing. Six groups of 15 *Ss* were run, three groups in each condition. The *Ss* were primarily college students.

Results. The results for the recall-first condition, the recognition-first condition, and both combined, are presented in Fig. 4, 5, and 6 respectively. For both conditions, recall for the critical figure was as good for the groups given *B* as for those given the unchanged *A*. Recall for the groups given *C*, on the other hand, was clearly poorer. Paralleling these results for recall were those for recognition (responses of 'same' or 'similar'). There was only a slight decline for the groups given *B*, whereas there was a sharp decline for those tiven *C*. Because these trends were so similar for the two conditions, and because the small number of *Ss* per group otherwise made statistical comparison difficult, the following analyses were based on the data for the two conditions combined (Fig. 6).[13] It can be seen that recall for the critical item was the same for the groups given *A* and *B* and that recall of the noncritical items was at about the same level for these two groups. For the groups given *C*, recall for the critical item dropped precipitously, thereby departing from the level of recall for the noncritical items, which remained at the same level for the groups given *C* as for those given *A* and *B*. The difference between recall of the critical item for the groups given *C* and those given *A* was significant at the .01 level. Thus a

[13] There appear to be two differences between the results of the recall-first condition and the recognition-first condition. One is Group *C*'s lower values of both recall and recognition for the critical item when recognition was first. This was probably the result of a somewhat different set when recognition was not tested in the context of recall as it was in the recall-first condition and in Experiment II (see that section; see also n. 5). The other difference between the two conditions is in the absolute level of recall among the various groups. For recall-first, the three groups seemed to differ in ability as reflected by the scores for noncritical items. But it is clear from a comparison of critical and noncritical items that only for the *C* group was there a decline in recall of the critical item.

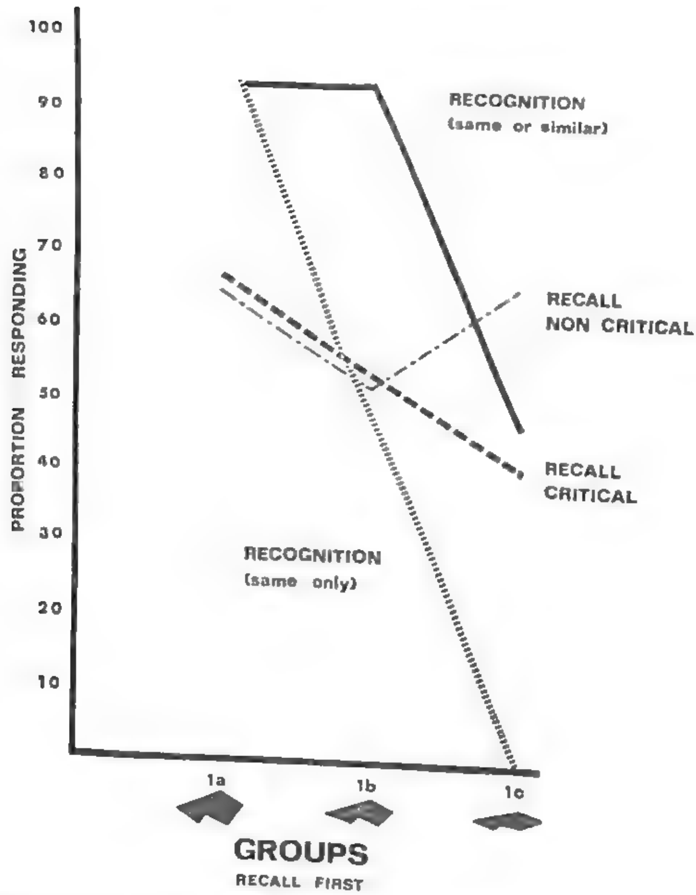


Fig. 4. Recall (Critical and Noncritical) and Recognition (Same or Similar and Same Only): Experiment III. Recall-First Condition

gradient was obtained. The decline in recall of the critical item for the groups given C is fully accounted for by the Ss' failure to recognize C, for of those cases in the three groups where C was recognized as the 'same' or 'similar,' the percentages of recall of the critical item were 62%, 68% and 67% respectively. Hence it is again fair to conclude that the gradient of recall obtained was produced by the nonrecognition of the altered figure. As in Experiment II, *where there was no recognition, there was no recall*; here, these failures of recall were always cases of no response rather than incorrect guesses. (It may be noted that the transformation of B was apparently not enough to lead Ss to the phenomenal impression that it was much different from A; they all regarded it as at least similar.)

Once again, the plot of 'same' recognition responses revealed a sharp gradient. Therefore it is quite clear that many Ss who recog-

nized B and C as 'similar,' and who may also have recalled the associated word, were nevertheless clearly aware that the figure was not identical with the one shown in training. The recognition of noncritical figures was again good. Of 450 test presentations, there were only 17 cases of nonrecognition. None of these 17 were cases where recall occurred, and all were no-response errors.

Discussion

One might take the position that the experiments reported are banal because it is obvious on intuitive and logical grounds that recall presupposes recognition in paired-associates learning. Intuitively, the act of recall may be said to imply giving an association to an item *one has encountered previously*. In terms of the logic outlined in the first paragraphs of this paper, recognition of A

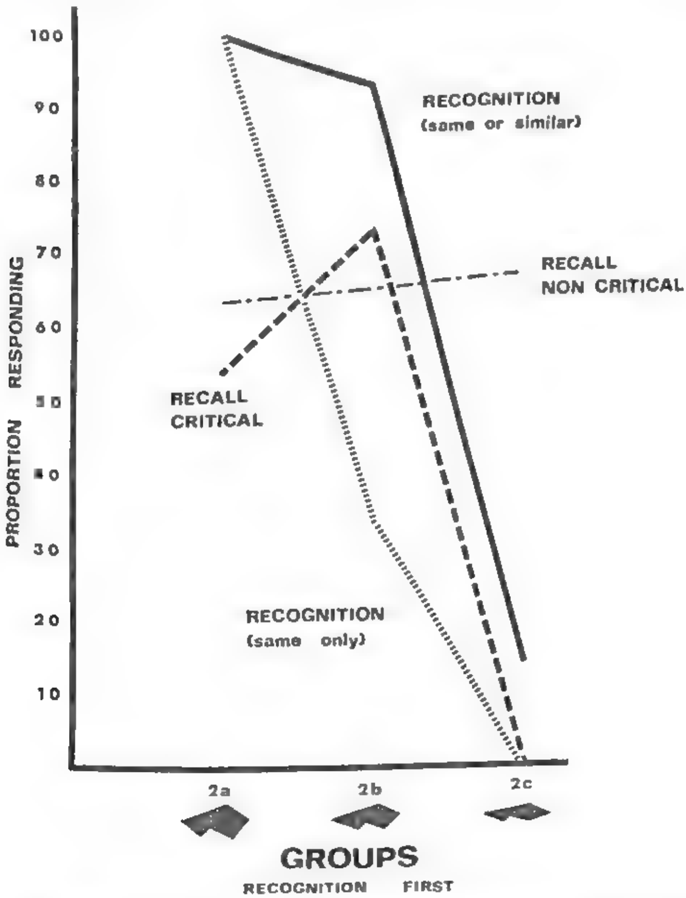


Fig. 5. Recall (Critical and Noncritical) and Recognition (Same or Similar and Same Only): Experiment III, Recognition-First Condition

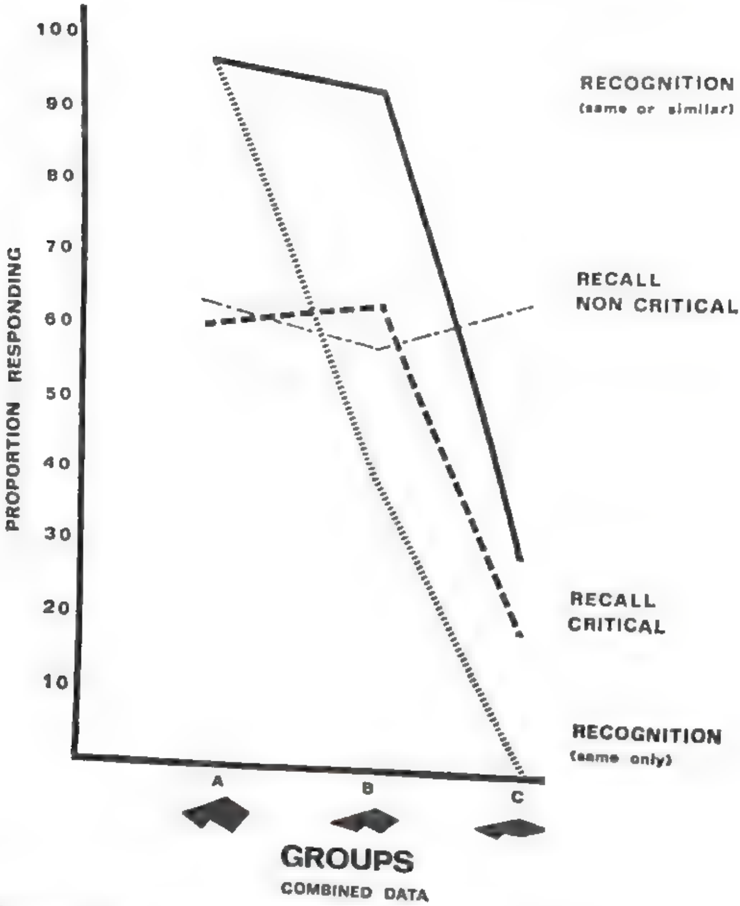


FIG. 6. RECALL (CRITICAL AND NONCRITICAL) AND RECOGNITION (SAME OR SIMILAR AND SAME ONLY): EXPERIMENT III, COMBINED DATA

should be expected to be a prerequisite for recall of an associated B. If this is so, recognition or the absence of it is the obvious source of 'generalization' or the absence of it.

Yet one does not find any linkage of recognition and recall in the literature (except for the few instances where Gestalt psychologists cite the argument of Höffding).[14] They are considered independent ways of testing retention. One reason for this is that when recognition is tested, the task is typically different than when recall is tested. In the former case, *S* learns a list of single items, generally in randomly varied order; in the latter case, he learns associations, either serially or in pairs. No attention was ever paid to

[14] For an analysis of the implication of Höffding's argument, see also Rock, in *Theories of Mind*, 1962 (see n. 1).

the obvious fact that in the associative situation the recognition of the constituent items would be occurring (or failing to occur).

Another explanation of the failure to see the importance of the relationship between recognition and recall is the fact that theorizing about learning has been based upon association psychology. If all learning is associative, then it is not clear what recognition *is*, and with very little exception, one does not find discussion of this problem in the literature. A few psychologists have tried to treat recognition as the emitting of a prior learned response—the "recognition response"—an hypothesis which is very hard to understand. It would seem to require the assumption that during learning *S* associates an item with the response 'I recognize this.' Such an assumption is clearly absurd. Yet if we don't assume this, how can we account for a response never associated with that item? If, on the other hand, some learning is not associative but rather consists of the depositing of a memory trace of a sensory event, recognition is perfectly understandable. In other words, one can learn *items* in the sense that one can later recognize or recall them.[15] The present authors submit that the associative bias has prevented psychologists from viewing stimulus 'generalization' in terms of the reaction to the stimulus term itself rather than in terms of the connection between it and a response. Although psychologists have begun to use recognition as an index of 'generalization,' *e.g.* in areas such as perceptual learning and semantic generalization, it seems implicit that recognition is considered an associated response of some kind and, therefore, to pose no special problems.[16]

There are, of course, also difficulties with the presently proposed theory. If *S* is unaware that a test stimulus is different from the training stimulus, 'generalization' may well then be expected. If, however, *S* is aware of the difference, but also aware of the similarity, then it becomes somewhat more problematical as to what he will do insofar as responding with the association is concerned. We have argued that the similarity suffices to tap the appropriate

[15] I. Rock and J. Ceraso, Toward a cognitive theory of associative learning, in C. Scheerer (ed.), *Cognition: Theory, Research, Promise*, 1964. We include recall because item-learning can also be tested by unaided or free recall and it is not at all proven that free recall entails any associative process. See S. E. Asch and S. M. Ebenholtz. The process of free recall: Evidence for non-associative factors in acquisition and retention. *J. Psychol.*, 54, 1962, 3–31.

[16] J. J. Gibson and E. J. Gibson, Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 62, 1955, 32–41; B. J. Underwood, False recognition produced by implicit verbal responses, *J. exp. Psychol.*, 70, 1965, 122–129.

trace (A) and, once that occurs, that the associated B will come to mind as often as when A remains identical. But one might also argue that even if it does come to mind, *S* must decide whether or not the changed stimulus 'calls for' the response. In the experiments reported in this paper, one would imagine an *S* to feel that a 'similar' stimulus *represented* one of the training figures and that a 'different' stimulus did not. The data suggest that the *Ss* did behave in accordance with this understanding. In the circle study, 'same' recognition responses were preceded by correct recall 78% of the time and 'similar' recognition responses were preceded by correct recall 78% of the time; in the ellipse study, the corresponding values were 65% and 64% respectively; in the main part of Experiment II, on orientation, the corresponding values were 45% and 50% respectively; and in Experiment III the corresponding values were 61% and 72%. Hence it is clear that awareness of 'similarity with a difference' led to correct recall as frequently as did the experience of identity.

Although the data of Experiments II and III were analyzed in terms of whether or not a gradient was obtained, the recognition theory presented is clearly an all-or-none theory: If recognition occurs there will be no response decrement; if it fails to occur, there will be no response at all. Hence a gradient, if it is obtained, is an artifact of a measure based on probability of response. The probability of nonrecognition increases as the phenomenal difference of the stimulus increases and more *Ss* fail to recognize it. If the same *S* could be tested many times without losing naïveté, then a gradient could be obtained even for one *S*. In either case the gradient would not represent varying degrees of habit strength. How then can the recognition theory explain a gradient generated by response measures such as amplitude, latency, or rate of response? We would suggest the following explanation. Stimuli which are identical to or not discriminably different from the training stimulus will, of course, yield responses of unchanged amplitude, rate, or latency. Stimuli which are perceived as similar yet discriminated as different from the training stimulus will create uncertainty in the mind of the *S*, who must decide whether or not the stimulus is the signal for the unconditioned stimulus, *i.e.* does represent the training stimulus. This state of uncertainty could easily lead to increased latency or decreased rate or amplitude of response. Data from another method that has been used in studying the problem of 'generalization' may have some bearing on this issue. According to the

method introduced by Lashley and Wade, the animal, after training, is required to discriminate the training stimulus from one similar to it.[17] In their study, the training stimulus was not preferred, from which they concluded there was no gradient, that 'generalization' was total because the animal had not attended to all specific features of the training stimulus. However, other experimenters using this method have demonstrated that the animals do show a preference for the training stimulus. MacCaslin, Wodinsky, and Bitterman have shown that whether or not a preference is manifested depends upon the difficulty of the original training.[18] If it is sufficiently difficult, the animal is forced to attend to the training stimulus, with the result that various of its features (*e.g.* size, shape, color) are noted. That being the case, the animal will be aware of the difference between the test stimuli. Restating this in terms of our present hypothesis, both test stimuli are recognized as similar to the training stimulus, but one is experienced as identical, the other as different in some respect. Naturally, this method will yield a gradient just as did the present experiments when plotted by the responses of 'same' only.

According to the explanation of stimulus 'generalization' suggested here, as in that advanced by Lashley and Wade and by Razran, 'generalization' occurs at the time of the test and not during learning.[19] It is denied that associations are formed during learning between potential stimuli—which lie along a continuum of similarity to the stimulus used in training—and the associated response. Often, such a continuum would have no existential reality for the organism and, even if it did, it does not follow that the neural representatives of stimuli which lie on it would in any way be aroused or come to mind during training unless *several* stimuli along the continuum were used during training. Instead, as noted above, 'generalization' can be understood in terms of a 'tolerance' in the process of trace arousal based on similarity at the time of the test.

Summary

If the mere perception of a stimulus object leaves behind a memory trace, and if the trace adequately represents the stimulus object, then the learning of items is often not an associative process.

[17] Lashley and Wade, *loc. cit.*

[18] E. F. MacCaslin, J. Wodinsky, and M. E. Bitterman, Stimulus generalization as a function of prior training, this JOURNAL, 65, 1952, 1–15.

[19] Lashley and Wade, *loc. cit.*; G. Razran, Stimulus generalization of conditioned responses, *Psychol. Bull.*, 46, 1949, 337–365.

Rather, if recognition is defined as the arousal of the trace when the stimulus object is encountered again, and if this arousal is based on similarity, then recognition should occur even if the stimulus is to some extent different. Thus, it is this recognition of similar, but not identical, stimuli that may be the essence of 'generalization.' (Research with animals does not directly test recognition, but the elicitation of a conditioned response by a similar stimulus reflects the fact that the stimulus is recognized.) That is, the first step in associative recall may be the arousal of the trace of the stimulus (recognition). In line with this theory, the present experiments presented evidence that in associative learning, recall presupposes recognition. The probability of recall of the item paired with an altered stimulus was shown to be a function of whether or not that stimulus was recognized, and recognition was shown to occur even when *S* was aware the stimulus was not identical to the one learned, so that 'generalization' is *not* equivalent to failure of discrimination.

ON PRODUCING THE MEANING IN SENTENCES

By DONALD G. MACKAY AND ROBERT W. BOWMAN, JR.,
University of California, Los Angeles

Clarification of the roles of syntax, semantics, and phonology in speech production was the main goal of this paper. One specific question was whether practice in producing the meaning of a sentence facilitates speech rate independently of syntactic or articulatory factors. To answer this question, bilinguals were asked to produce translations of sentences they had practiced in their other language. If these translations were easier to produce than nontranslations, with appropriate controls for syntax and articulatory factors, we could conclude that practice at a semantic level facilitates the rate of speech. Other questions were: Does practice with the meaning of a sentence reduce the probability of errors in speech production? Are there different psychological types of meaning, as Katz and Fodor postulated?[1] Can speech production be viewed as an analytic process, whereby a general propositional meaning is analyzed into a set of specific meanings before being transformed into words?

STUDY I

PRACTICE, BILINGUAL TRANSFER, AND RATE OF SPEECH

Hypothesis 1: On Practice Effects in the Production of Sentences

MacKay has elsewhere demonstrated that bilinguals are able to speak faster in the language with which they are more familiar.[2] Here, familiarity was varied by having bilingual *Ss* repeatedly produce a single sentence in one language in a short period of time. If

* Received for publication August 26, 1968. The work reported was supported by UCLA Grant 2428 and by funds of the Department of Psychology, UCLA, that were provided by Dr. F. N. Jones. The authors wish to thank Drs. Donald Lindsley, Donald Walter, George Mount, Edward Carterette, and Barry Collins, for generously lending various pieces of equipment for the present study, and Drs. Tom Trabasso and Eric Holman, for their excellent suggestions.

[1] J. Katz and J. Fodor, The structure of a semantic theory, *Language*, 39, 1963, 170–210.

[2] D. G. MacKay, Of age and experience: Factors in the control of speech under delayed auditory feedback, forthcoming.

short-term and long-term practice have similar effects, the time to produce a sentence should decrease with repetition even though the *S* is speaking at his maximum rate throughout.

Procedure. Twelve German-English bilinguals were paid for participation (in this and the experiments to follow). All were students at UCLA. Eleven of the *S*s were native German speakers who had learned English in high school and had lived in America for at least a year. The remaining *S* was an American who had spent several years in Germany. All of the *S*s were fluent in both languages, having spoken their second language for an average of six years. Their mean age was 24 years, with a range of ages from 18 to 38. These *S*s were presented with a sentence typed on an index card, and they were then instructed:

This is a study in speech production, and we are interested in knowing how fast sentences can be read in different languages. I'll give you a set of sentences typed on cards like these, one card at a time. [*E* then gave examples of sentences *S* would be reading.] Each sentence is to be read as fast as possible. Don't worry about making errors, but don't omit words. I'll tell you beforehand whether the sentence is in English or in German. I'll give you the card face down, and when I tap the table, turn the card over and read the sentence. There will be 12 repetitions of each sentence, and there are 12 sentences in all.

Half the sentences were in German and half were in English. They were obtained from German literary works with available English translations.[3] The sentences averaged 15 syllables in length, with a maximum number of 16 and a minimum of 14 syllables. The trials were recorded on a Sony-Matic (TC-105) tape recorder. Two judges independently determined the times for producing the sentences, by stopwatch; they did not know whether a sentence they were timing was a translation or a nontranslation.

Results. The average times to produce the sentences show that the rate of speech increased for successive repetitions of a sentence. (Fig. 1, left section) The difference in production time for the first four and last four repetitions was significant at the .01 level, using a two-tailed, signed-ranks test with *S*s as the unit of analysis. Clearly, the maximum rate of speech was not fixed for a given individual but varied with his practice or familiarity with the material he was producing. This increase in speech rate may, however, have been due to practice with the phonological, syntactic, or semantic levels of the sentences, so the next point of interest became one of determining whether practice in producing the *meaning* of a sentence reduces the time required to produce that sentence.

[3] F. Durrenmatt, *Der Besuch Der Alten Dame*, 1957; *The Visit*, P. Bowles (trans.), 1962. F. Kafka, *Der Prozess*, 1918; *The Trial*, W. and E. Muir (trans.), 1948. E. Remarque, *Im Westen Nichts Neus*, 1929; *All Quiet on the Western Front*, A. Wheen (trans.), 1929.

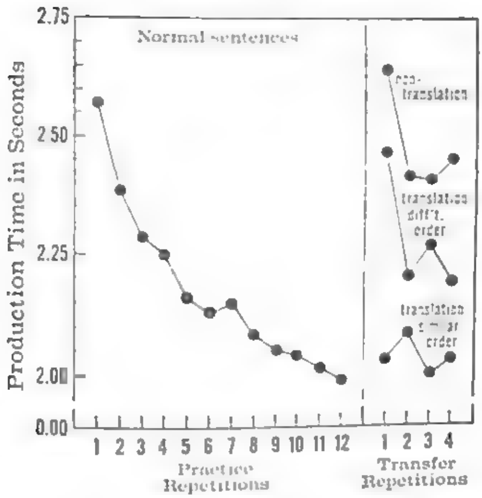


FIG. 1. MEAN TIMES FOR PRODUCING NORMAL SENTENCES DURING PRACTICE AND IN TRANSFER AFTER PRACTICE: HYPOTHESES 1, 2, AND 3

(The mean times were averaged over *Ss*. The increase in maximal speech rate resulting from practice—as shown at the left—transfers across languages—as shown at the right—for translations, the greatest transfer occurring for translations with a word order similar to that of the practice sentences.)

Hypothesis 2: On Semantic Facilitation Across Languages

A transfer paradigm was employed to test the hypothesis that familiarity with the meaning of a sentence increases the maximal speech rate. The *Ss* were again instructed to repeat a sentence 12 times, at their maximal rate. This procedure is described as the practice phase. The *Ss* were then given the following instructions in what is herein called the transfer phase:

> After the practice sentence you will be given a second sentence which you are to read as fast as possible, four times in rapid succession. You will be told whether the sentence will be in English or German.

The *Ss* were given two practice trials with practice and transfer sentences to ensure that they understood instructions. When the transfer sentence was in German, the practice sentence was in English, and vice versa. On half the trials, the transfer sentence was a translation of the sentence *S* had practiced. For example, if *S* practiced the sentence, *Das konnte jeder merken, der ihre fahle Haut ansah*, then the sentence in his transfer phase was either *Anyone who looked at his sallow skin could see that*, or *For no man is the earth so meaningful as a soldier*. The sentence beginning *Anyone who* is a translation of the practice sentence, with the same meaning but different word order and speech sounds, but the sen-

tence beginning *For no man* bears no relation to the practice sentence in phonology, syntax, or meaning. According to the semantic-facilitation hypothesis, which suggests that practice in producing the meaning of sentences increases rate of speech, therefore, the sentence with the same meaning as the practice sentence (*Anyone who*, here) should be rapidly produced, but production of the sentence whose meaning is irrelevant to the practice sentence (*For no man*, here) should not be facilitated.

Procedure. The order of the practice sentences was randomized for each *S*. In order to equate the difficulty of the transfer and practice sentences, the transfer sentences for half the *Ss* served as the practice sentences for the remaining *Ss*. The difficulty of the translated and nontranslated sentences in the transfer phase was equated by making each sentence in the transfer phase a translation for half the *Ss* and a nontranslation for the other half of the *Ss*. Thus, for different *Ss*, each sentence was used an equal number of times both as a translation and as a nontranslation. A further control for sentence difficulty was to have two pilot *Ss* read each sentence at their maximum rate. Their mean rate and *SD* were then calculated for the 36 sentences, and sentences which were produced in much longer or much less time than the mean were replaced by other sentences. There were thus selected 24 sentences in German, with their English translations, making 48 sentences altogether. (Other examples of these sentences appear in Table I, under "With Different Orders.") Each *S* received six sentences in each language in the practice phase.

Results. The results are shown in the panel at right in Fig. 1. One question was whether sentence repetition resulted in fatigue. The rate of speech for the first four practice sentences was slightly faster than for the nontranslation test sentences. However this difference was not significant at the .20 level using a two-tailed, signed-ranks test. Thus, if there was a fatigue effect it was very small. The translations were produced faster than the nontranslations in the transfer phase, as predicted by the semantic-facilitation hypothesis. This difference was significant at the .05 level (using a two-tailed, signed-ranks test). It is obvious that this difference was not the result of sentence difficulty, since the same sentences served as both translations and nontranslations for different *Ss*. We can only conclude that the increase in speech rate for the translations was due to practice in producing the meaning of a sentence, since the translation sentences differed from the practice sentences in all respects but semantics.

Hypothesis 3: On Word-Order Effects

The translations in the test of Hypothesis 2 varied the word order of the original sentence, as translations often do. Since word

TABLE I

Examples of Normal Sentences

Translations with similar orders

Besten er plündert uns jetzt als nach dem Besuch der Dame.	Best he plunder us now and not after the lady's visit.
In einer Ecke des Zimmers standen drei junge Leute.	In one of the corners of the room were standing three young men.
Die Häuser auf unserer Seite sind geräumt worden.	The houses which are on our side have been abandoned.
Das Mädchen auf der Bretterwant ist für uns ein Wunder.	The girl on the poster is for us a complete wonder.
Die andern setzen sich und lehnen die Rücken an das Gitter.	The others sit down and lean their backs up against the fence.

Translations with different orders

Er hätte geradewegs in sein Zimmer gehen können.	It had been his intention to go directly to his room.
Ich habe an einem Morgen vierzehnmal sein Bett gebaut.	I have rearranged his bed fourteen times in one morning.
Auch ist dieser Blick in den Spiegel wirklich angenehm.	This moment before the mirror is truly a pleasant one.
Was würden wir geben, wenn wir zu ihnen zurück könnten!	What we would not give to be able to go back to it!
Vor vierzehn Tagen mussten wir nach vorn, um abzulösen.	Fourteen days ago we had to go and relieve the front line.

order may be an important determinant of the practice effect in itself, however, transfer for translations of practice sentences having the same word order was then compared with transfer for translations having a different word order. For example, consider the pair, *Für niemand ist die Erde so viel wie für den Soldaten* and *For no one is the earth so meaningful as for a soldier.* Here the English translation retains the exact word order of the original German sentence. If word order plays a role in practicing sentences, the rate in the transfer phase for such a syntactically identical translation should be greater than for a translation with altered word order, such as *The earth is more meaningful to a soldier than anyone else.* (See Table I for other examples.) Instructions, *S*s, and procedure were the same as in the investigation of Hypothesis 2. The order of presentation of the two types of translations was randomized for each *S*.

Results. Fig. 1 shows that transfer for the translations in the same word order was greater than for the translations in a different word

order. This difference was significant at the .05 level using a two-tailed Wilcoxon test. One explanation of this finding is that bilinguals may associate a series of words in one language with its translation in their other language. In some sense, then, our *Ss* might already have practiced the translations in the transfer phase while producing the practice sentence. Since this low-level explanation has not been conclusively rejected in other studies of bilingual transfer,[4] it was tested with the following experiment.

Hypothesis 4: On Facilitation at the Word-Level—An Alternative Explanation of the Practice Effect

Katz and Fodor pointed out that a semantic theory which considers only denotative or "dictionary" meanings is inadequate, since it excludes the additional meaning gained by analyzing the contextual relationships among the words of the sentence.[5] Thus, two levels of semantics can be distinguished: the lexical or word-dictionary level and the sentential level. Meaning at the sentential level depends on the contextual interaction of the dictionary meanings of words. For example, in the sentence *He fed the information into the computer's memory store*, the lexical item *store* cannot mean *delicatessen* in that context. If the words of a sentence are scrambled, however, this contextual or sentential meaning is destroyed or severely disrupted even when the lexical meaning is still intact. Use of such scrambled sentences could therefore help determine whether semantic facilitation between languages occurs at the lexical or sentential level of meaning. If significant transfer were found for translations of scrambled strings, it could be argued that the bilingual transfer found in this and other studies was based on cross-linguistic association made at the word level while *Ss* read the practice sentences.

Procedure. The same 12 *Ss* were instructed to produce two types of sentences at their maximal rate, as if all the sentences were normal ones in German (or English respectively). One type was identical to that used in the test of Hypothesis 1. *e.g. Dann packte auch ihn wie einst den Grossvater die Wanderlust* (*Then the wanderlust seized him as it once had his grandfather*). For the other type, the order of the words of the sentences was scrambled to give strings such as *Den auch wie packte dann Grossvater einst Wanderlus ihn die* (*The also as seized then grandfather once wanderlust him the*). (See Table II for other examples of scrambled sentences.) For the transfer phase,

[4] P. Kolers, Bilingualism and information processing, *Sci. Amer.*, March, 1968; Interlingual facilitation of short-term memory, *J. verb. Learn. verb. Behav.*, 5, 1966, 314–139.
[5] Katz and Fodor, *loc. cit.*

half of the scrambled sentences were translated as shown in the parentheses. The other half were nontranslations, such as *Some airplane was the pilot tight for in spiral the engaged.* Each *S* produced eight scrambled pairs. The scrambled sentence in the transfer phase was a translation of the practice sentence for half the *Ss* and irrelevant to the practice sentence for the other half of the *Ss*. Comparison of translations and nontranslations was expected to determine whether facilitation was taking place at the lexical level.

Results. The time to produce the sentences was determined by stopwatch as before. The effect of practicing scrambled sentences is shown in the left panel of Fig. 2. Practice increased the rate for the scrambled sentences in the same way as for the normal sentences. The average rate in the last four repetitions was significantly faster than that in the first four, a difference significant at the .01 level determined by a two-tailed, signed-ranks test. The transfer phase is shown at the right of Fig. 2. No facilitation was found for the translations of the scrambled sentences. The difference between the translations and the nontranslations in the transfer phase was not significant (.05 level, signed-ranks test). Practice did not facilitate the production of translations of the scrambled sentences, as it did the normal sentences. In fact, the translations of the scrambled sentences were produced at a slightly slower rate than the nontranslations. Both the direction of this difference and the peculiar shape of the function for scrambled sentences were unexpected. At present, we are unable to explain this aspect of the data.

Comparing the scrambled with the normal sentences, we can conclude that practice with the meanings of individual words does not facilitate the production of speech whereas practice with the patterned meaning of sentences does—a finding reminiscent of the Gestalt principle that the whole is greater than the sum of its parts. More specifically, the lack of facilitation for the scrambled sentences

TABLE II

EXAMPLES OF SCRAMBLED SENTENCES

Damit ich Angelegenheit erledigt ist die glaube.
Therefore I event preceding settled is the do believe.

Sie Hundelauferei wissen diese nicht verboten ist.
You dogs running about to know this not forbidden is.

Rief Herr er gut sie Sudermann hat gehalten gut sich sehr.
Called Mister he good she Sudermann has stood well it very.

Tau Blumen Amsel den auf Gartens eine sang auf lag des.
Dew flowers blackbird the on garden a sang on lay of the.

Wunderbar wir die zusammen all war waren es Tage.
Wonderful we used those together all was were it days.

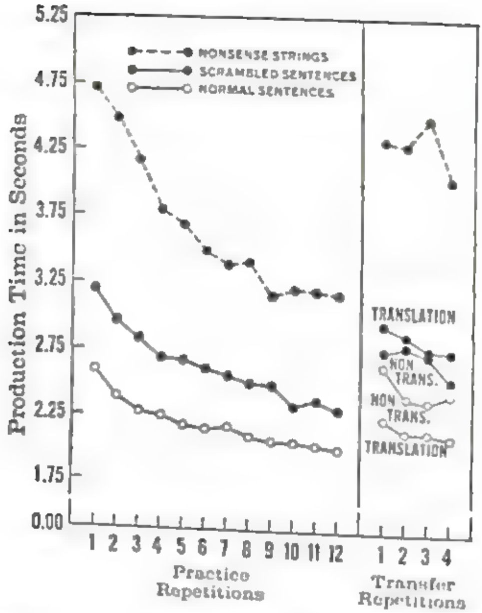


FIG. 2. MEAN TIMES FOR PRODUCING NORMAL AND SCRAMBLED SENTENCES AND NONSENSE STRINGS DURING PRACTICE AND IN TRANSFER AFTER PRACTICE: HYPOTHESES 4 AND 5

(The mean times were averaged over *Ss*. The increase in maximal speech rate resulting from practice—as shown at the left—transfers across languages—as shown at the right—only for normal sentences, *i.e.* only as a result of sentential practice. For translations of normal sentences, word order differed from the practice sentences; for those of scrambled sentences it did not.)

suggests (1) that semantic facilitation operates at the sentential level of meaning rather than the dictionary level and that, in fact, semantic transfer may reflect practice in integrating meanings of individual words into the context of the sentence as a whole; (2) that the bilinguals were not translating the words to themselves as they read the practice sentences; and (3) that the semantic facilitation effects shown earlier were not due to an intrinsic association between lexical items in the stored representation of the two languages of bilinguals. This finding also suggests that distinctive features of meaning must be relative rather than absolute. If semantic effects cannot be found at the word level, we are unlikely to find anything resembling a word dictionary in the head. If the psychological meaning of a word depends on its sentential context, it is doubtful that distinctive features of meaning for a word can be found which apply to all possible contexts at all possible times. A further objection to absolute rather than relative distinctive features is that psychological meaning simply cannot be reduced to a common de-

nominator of what is present in all contexts in which a word is used, as Nida and Wittgenstein have pointed out.[6]

Hypothesis 5: On Phonological Practice

In order to test for effects of practice at the phonological level, syntactic and semantic features had to be eliminated from the sentences. Syntactic features were eliminated by scrambling normal sentences to give strings like *Rites man such never primitive witnessed before had modern*. Semantic features were then eliminated by inverting or substituting letters in the words of the strings. The changes were such as to maintain the normal phonological pattern of English or German, depending on the language of the original sentence. Examples of the nature of the changes can be found by comparing the sentence *Rickes mun sach naiver primodave wotnezzed befer hap medorn*, with that above. The changed strings were thus syntactically, sententially, and lexically meaningless but carried a normal phonological structure. Pilot *Ss* (American and German) read the strings to ensure that they were meaningless but pronounceable in the respective languages. (Additional examples appear in Table III.) The *Ss* were instructed to produce a total of eight such strings 12 times at their maximum rate as if they were normal sentences in German or English respectively.

TABLE III

EXAMPLES OF NONSENSE STRINGS

German

Gleuche meis sehen alber won Echnung ürger Geblandenheit.
Em gübess sumisch Redelt, wochst angedundeln neschaugen.
Niebe sulehen Gernisch dirn urn dam zornum angefegen.
Schlämpfen zund Herfelersten reiblend win angekarpischte.
Gie hernte mis ehnet Verglubnet wolchgatten gnäliben.

English

Lipe thost sperths pavering noarly thas anerlous emkerment.
Selpferth poaled entsing bort thamp rebunole hest thenticious.
Helted marthole grust severlate en foalt mesker raltingly.
Ortments sebning glabually therp samful marding tormous.
Heffing lornily thort nirlish goy ound pledganous soful.

Results. As the left section of Fig. 2 shows, practice with these nonsense strings increased the rate of speech in the same manner as did practice with the semantically and syntactically meaningful

[6] E. A. Nida, *Morphology: The Descriptive Analysis of Words*, 1946; L. Wittgenstein, *Philosophische Untersuchungen*, G. E. M. Anscombe (trans.), 1953.

sentences. Again, the average of the last four repetitions was faster than that of the first four, a difference significant at the .01 level using a two-tailed, signed-ranks test. We can therefore conclude that practice in producing the phonology of sentences improves the rate of speech. In order to test whether this practice effect transferred across languages, the *Ss* then produced nonsense in their other language as in the tests of Hypotheses 2-4. No transfer was found. As may be seen at the right of Fig. 2, the rate for the first four repetitions in the transfer phase was no faster than that for the first four repetitions in the practice phase. This finding conclusively rules out the possibility of a phonological interpretation of the previously reported semantic and syntactic facilitation. Finally, it is interesting that reading the nonsense strings was so much slower than reading the scrambled sentences and normal sentences (see Fig. 2, at the left). This can be interpreted to reflect the practice the *Ss* had had with the syntax, phonology, and sentential meaning of the normal sentences. Study II then proceeded to test the generality of the semantic and syntactic facilitation shown in Study I.

Study II

Practice, Bilingual Transfer, and Errors in Speech

Hypothesis 6: On Semantic Facilitation Under Delayed Auditory Feedback

Hypothesis 6 postulated that practice in producing the meaning of a sentence reduces the probability of error in speech production. Delayed auditory feedback (DAF) served as a technique for inducing errors in speech, since normal *Ss* experience a temporary speech disorder when their auditory feedback is delayed for a fraction of a second.[7] They will stutter, prolong and omit speech sounds, and produce phonemes which are not part of any language with which they are familiar.[8] MacKay has also found that the speech of bilinguals under DAF is disrupted less when they are speaking in the language with which they are more familiar.[9] Thus, if short-term and long-term practice have similar effects, practice with the meaning of a sentence should reduce errors in speech under DAF.

Procedure. As in Study I, *Ss* repeated a sentence 12 times with normal auditory feedback. The *Ss* then repeated four times a transfer sentence which was

[7] B. S. Lee, Artificial stutter, *J. speech hear. Dis.*, 16, 1951, 53-55.
[8] G. Fairbanks and N. Guttman, Effects of delayed auditory feedback upon articulation, *J. speech hear. Res.*, 1, 1958, 12-22.
[9] MacKay, *loc. cit.*

a translation of the practiced sentence on half the trials. In the present study, however, Ss read this second sentence under DAF. The 12 Ss (the same as those in Study I) were instructed:

We want to know how fast people can speak different languages under DAF. DAF is a method for inducing stuttering in people who don't normally have this problem, and you'll be wearing headphones and speaking into a microphone. Otherwise, everything is the same as in the first experiment.

E then reviewed the procedure to be followed and gave *S* several sentences to be read under DAF. The apparatus used to delay feedback was an Echo-Vox variable-feedback system (Kay Electric Company), in conjunction with a set of Permoflux (PDR 600) stereo headphones and a Monarch TM-18 microphone. The microphone was connected to a stereo amplifier (Knight KN-721). Speech conducted to the headphones in this manner produced a maximal sound pressure level of 95 db. The delay in feedback was set at .2 sec. The microphone was adjusted so that it was approximately 6 in. from the lips of *S*. Each trial was recorded on tape. Two experienced judges independently determined the time to produce each sentence and the number of stutters under DAF. A stutter was defined as the repetition of speech sounds equal to or less than a syllable in length, as in twit-itering. Stuttering was the only error analyzed.[10]

Results. The probabilities of stuttering for the translations and nontranslations appear in Table IV. The Ss stuttered less for translations than for nontranslations, a difference significant at the .05 level with a two-tailed, signed-ranks test. This suggests that practice or familiarity with the meaning of a sentence facilitates the control of speech under DAF, since the only factor in common between the practice and transfer sentences was the semantics. Thus, practice in producing the meaning of sentences has an effect on the control of speech similar to its effect on speech rate as shown in Study I. Moreover, we can conclude that the increased rate and control of speech were independent effects of practice, in view of repeated demonstrations that increasing the rate of speech increases rather than decreases DAF interference.[11]

Other results further characterized the role of meaning in speech production. (*1*) *Semantic errors.* Semantic intrusions occurred re-

[10] See MacKay, *loc. cit.*, for justification.

[11] J. I. Beaumont and B. M. Foss, Individual differences in reacting to delayed auditory feedback, *Brit. J. Psychol.*, 248, 1957, 85–89; S. Fillenbaum and R. Wieson, Contextual constraints and disruption in reading with delayed auditory feedback, *J. acoust. Soc. Amer.*, 33, 1961, 1800–1801; N. Guttman, Experimental studies of the speech control system, unpublished doctoral dissertation, University of Illinois, 1954; E. Kodman, Controlled reading rate under delayed speech feedback, *J. aud. Res.*, 1, 1967, 186–193; D. G. MacKay, Similarities between pathological stuttering and stuttering induced by delayed auditory feedback, forthcoming; D. G. MacKay, Metamorphosis of a critical interval: Age-linked changes in the delay in auditory feedback producing maximal disruption with speech, *J. acoust. Soc. Amer.*, 43, 1968, 811–821.

TABLE IV
PROBABILITIES OF STUTTERING TRANSLATED AND NONTRANSLATED NORMAL AND SCRAMBLED SENTENCES: HYPOTHESES 6 AND 7

	Probability of stuttering (per syllable)	
	Translations	Nontranslations
Normal sentences	.017	.032
Scrambled sentences	.024	.022

Note: $N = 12$; 48 sentences per *S*.

peatedly. For example, one *S* misread the sentences *Doctor Kellerman said you should not let your head hang so* and *Doctor Kellerman said you should not let your Kopf hang so*, where *Kopf* in the second sentence is a semantic intrusion—a translation of *head* in the first sentence. Kolers encountered similar intrusions when bilinguals read stories composed of phrases that switched from one language to another.[12] These errors resemble those of the dyslexic, reported by Marshall and Newcombe, who would misread *uncle* as *nephew*, for example.[13] Errors of a similar sort also occur in normals who, like the dyslexic, immediately recognize their error when it is pointed out.[14] These cases suggest that reading may be a constructive process taking place at the semantic level in part and dependent only in part on the visual input. Further evidence for the thesis that meaning facilitates the motor production of speech can be found in the errors of certain aphasics. Kurt Goldstein reported a case of a French aphasic who was incapable of reading and producing (on command) meaningless words such as *féca* or *céfa*; however, he could produce these same phonemes when they were combined to form the meaningful French word *café*.[15] Clearly, semantic facilitation must be one of the components of a general theory of the physiology of speech production.

(2) *Spoonerisms*. The *Ss* also occasionally inverted or reordered speech sounds. For example, one *S* produced *pedot*, reversing the order of the phonemes in the word *depot*. This *S* made the same error in all four repetitions of the sentence. These Spoonerisms were rare, but have also been found by MacKay.[16] (3) *Consistency*

[12] P. Kolers, Reading and talking bilingually, this JOURNAL, 79, 1966, 375–376.

[13] J. Marshall and F. Newcombe, Syntactic and semantic errors in paralexia, *Neuropsychologia*, 4, 1966, 169–176.

[14] J. Morton pointed this out in a talk presented to the UCLA Phonetics Laboratory, November, 1967.

[15] K. Goldstein, *Language and Language Disturbances*, 1948.

[16] MacKay, Of age and experience (see n. 2).

within subjects. An *S* frequently made the same error when repeating a sentence. One *S*, for example, in producing the German word *Seite* (*side*), read *Seite-te* in all four repetitions of the sentence. He explained that he could predict where he would make errors in subsequent readings of a sentence but was nevertheless unable to avoid making them again. Bloodstein reports that individuals who normally stutter are consistent in this same manner—stuttering repeatedly on the same words in a passage.[17] This suggests that certain words are difficult for an individual to produce, in one case with DAF, and in the other with normal feedback. (*4*) *Consistency between subjects.* Different *Ss* frequently committed identical errors under DAF. For example, 75% of the *Ss* mispronounced the German word for *sniffed* (which is *schnupperte*) as *schnuppertete*. The evidence for such between-subject consistency in studies of pathological stuttering is not clear, although it appears that different pathological stutterers are not as consistent in the words on which they stutter as were the *Ss* in the present study.[18]

Hypothesis 7: On Facilitation at the Word-Level—A Further Consideration

The hypothesis of lexical facilitation was rejected in Study I in favor of the conclusion that semantic facilitation occurs at the sentential level. Hypothesis 7 tested whether this conclusion also holds for DAF stuttering. As before, the possibility of lexical facilitation under DAF was tested with scrambled sentences (see Table II). The 12 bilinguals practiced a scrambled string 12 times and then produced the scrambled transfer string four times under DAF. The scrambled transfer string was either a direct translation or a completely irrelevant string.

Results. Again, no transfer in rate of speech was found, and there was no difference in number of errors between the translated and nontranslated strings. The probability of stuttering in a translated string was as great as that of stuttering in a nontranslated string (Table IV). The fact that no difference was present in the amount of stuttering between the translated and nontranslated strings extends the results of Study I. The conclusion that semantic facilita-

[17] O. Bloodstein, The development of stuttering, *J. speech Dis.*, 25, 1960, 219 and 366.

[18] S. F. Brown, The loci of stutterings in the speech sequence, *J. speech Dis.*, 10, 1945, 181–192; J. Eisenson and E. Horowitz, The influence of propositionality on stuttering, *J. speech. Dis.*, 10, 1945, 193–197.

tion derives from the sentential rather than the lexical level is strengthened. But since this test did not control for rate of speech, the probability of stuttering in a scrambled string was not directly comparable to that of a normal sentence. The final study was concerned with this question.

Study III

Meaningless Material and DAF Interference

Hypothesis 8: On the Effect of Meaning Per Se

Eisenson maintains that the control of speech in individuals who stutter improves when spoken material is made less meaningful.[19] When stutterers repeat a sentence, they stutter significantly less, a phenomenon known as the adaptation effect.[20] Bloodstein has suggested that this adaptation effect may be due to a reduction in the meaningfulness (propositionality) of the repeated passage.[21] Thus, the effects of repeating a sentence in the present studies could be attributed to a reduction in meaning rather than to practice in producing the meaning of a sentence.

Hypothesis 6 demonstrated a phenomenon analogous to the adaptation effect in that repetition of a sentence reduced stuttering under DAF. In view of the similarities between DAF and pathological stuttering shown by MacKay,[22] the adaptation effect could be interpreted as an effect of practice. In Study III we wished to ensure, therefore, that the effect of repetition was due only to practice and not to the reduction of subjective meaning postulated by Bloodstein. Accordingly, the meaningfulness of sentences was varied. One set of sentences was complete nonsense, another set was meaningful at the lexical level (scrambled sentences), and the third set was completely meaningful (normal sentences). If meaning disturbs the control of speech under DAF, as an extension of Eisenson's and Bloodstein's theories would suggest, then the completely meaningless strings should be produced with fewer errors than the completely meaningful sentences.

Procedure. The *Ss* were 21 German-English bilinguals, students at UCLA. Each set of sentences was composed of 10 sentences adapted from the German Bible.[23] (For examples of all three sets, see Table V.) The *Ss* produced each

[19] J. Eisenson (ed.). *Stuttering: A Symposium*. 1958, 225–271.
[20] W. Johnson. *Stuttering in Children and Adults: Thirty Years of Research at the University of Iowa*. 1955.
[21] O. Bloodstein. Conditions under which stuttering is reduced or absent: A review of the literature. *J. speech. Dis.*, 14, 1949.
[22] MacKay. Similarities (see n. 11).
[23] E. Beck and M. Gabriele (trans.). *Grossdruck Bibel*. 1965.

sentence only once, with instructions to speak at their maximum rate for all three types of sentences. The order of the three kinds of sentences was randomized for each *S*. The DAF apparatus was the same as before, but the delay was .2 sec. for 10 of the sentences and .175 sec. for the other 10. The order of presenting the sentence sets and the delays was separately randomized for each *S*. The average sound-pressure level was 95 db.

Results. The rate of speech and probability of stuttering are shown in Table VI. The probability of stuttering was least for the normal sentences, greater for the scrambled sentences, and greatest for the nonsense strings. However, the rate for the nonsense sentences was

TABLE V

EXAMPLES OF SENTENCES: HYPOTHESIS 8

Normal sentences

Nicht vom Brot allein lebt der Mensch, sondern auch vom heiligen Wort.
Es ziemt sich für uns alle Gerechtigkeit zu erfüllen.
Und da rief Herodes die Magier Heimlich zu sich.
Aber schon ist die Axt an die Wurzel der Bäume gesetzt.
Bereitet den Weg des Herrn und macht gerade seine Pferde.

Scrambled sentences

Angesagt Schafe zu gegen des nun Tochter Thimnath.
In sind Fürst Magdiel der wie Erblaude ihren und Edomiter.
Maulpferde der da erfand die hütete Wüste ist in.
Oberstes mein geträumet ein wir Weinstock und zu hat.
Alles gethan zu Gott solches in sprach könnten und Pharao.

Nonsense strings

Brauder heinter zoben nocht ieglich meim Keiber Büdet ges.
Niebe Ufühl Sahn Goldschmuck brauke Mauet aust breide.
Baues Redel ihret Zimeis un, derselden ihrel Sam.
Schmüle stut dur Shorstädt lan Majal Finder Meram gabet.
Zornum Sauchallan dam hierete win Maueln ein.

much slower than that for the normal sentences. As a control for the rate of speech, the *Ss* were instructed to produce an additional set of normal German sentences at their normal rather than maximal rate of speech. The rate of speech and probability of stuttering in this condition are also shown in Table VI. Slowing down the rate of speech further reduced the probability of stuttering, ruling out a rate-of-speech interpretation of the differences in stuttering for the three types of sentences, and confirming other reports.[24] This result supports the interpretation that meaning facilitates the control of speech under DAF rather than further disrupting it. And more generally, it provides an additional control for the conclusion

[24] See n. 11 for references on speech rate. Note also the difference in overall probability of stuttering between this group and the group in Study II. This difference may be attributed in part to sentence difficulty, subject differences, and the increased amplitude of the feedback in Study III.

TABLE VI

PROBABILITIES OF STUTTERING AND RATES OF SPEECH FOR NORMAL, SCRAMBLED, AND NONSENSE GERMAN SENTENCES AND STRINGS: HYPOTHESIS 8

	Probability of stuttering (per syllable)	Rate of speech (seconds per syllable)
Normal sentences	.10	.241
Scrambled sentences	.80	.285
Nonsense sentences	.15	.318
Control: Normal sentences, normal rate	.06	.254

Note: $N = 21$; 20 sentences per *S*.

that practice with the meaning of a sentence directly facilitates the production of speech.

As yet, however, we cannot conclude that meaning per se facilitates speech production independently of practice with the production of meaning. Other factors besides meaningfulness also varied in the three sets of sentences in this experiment. The word sequence of the scrambled sentences was novel and unpracticed. Similarly, the syllable sequence of the nonsense strings was unusual and unpracticed. Further research is needed to determine whether meaningfulness per se facilitates speech production and to test the hypothesis that the adaptation effect in pathological stutterers is due to practice.

Finally, the present findings support the view of speech production as an analytic process (as Wundt suggested).[25] Practice at the semantic level can be viewed as facilitating the analysis of a general propositional meaning (*Gesamtverstellung,* in Wundt) into more specific meanings, which are transformed into words. However, the data also suggest a constructive or synthetic process in speech production, since semantic practice effects seem to reflect the integration of meanings of individual words to form the contextually dependent *Gesamtverstellung.*[26] Clearly, then, the present findings are also consistent with an analysis-by-synthesis model of speech production, as proposed by Stevens and Halle.[27]

SUMMARY

Three studies focused on the practice effect in speech production—an effect which can be defined as the increase in maximum rate of

[25] W. Wundt, *Essays: Die Sprache und das Denken,* 1909.
[26] H. Paul, *Principien der Sprachgeschichte,* 2nd ed., Engl. trans. by H. Strong, 1889.
[27] K. Stevens and M. Halle, Remarks on analysis by synthesis and distinctive features, in Wathen-Dunn (ed.), *Models for the Perception of Speech and Visual Form,* 1967, 88–103.

producing a sentence as a result of practice. This facilitating effect of practice was separately demonstrated for all three levels of speech production.

Study I. (a) The semantic level. Bilinguals practiced producing a sentence in one language and then produced either a translation or a nontranslation of this sentence in their other language. Translations were easier to produce than nontranslations even though the sounds and syntax of the test sentences were completely different from the practiced sentence. This cross-language transfer was shown to depend on the sentential or context-dependent level of meaning rather than on the word or context-independent level. When the bilinguals practiced scrambled sentences, which were meaningless at the sentential level, a practice effect was found, but this did not transfer to the production of a translation in their other language. *(b) The syntactic level.* The bilinguals practiced a sentence as before and then produced either a translation which maintained the same word order or one with dissimilar order. Facilitation was greater for translations with the same word order than for translations with different word order. *(c) The phonological level.* The bilinguals practiced sentences which were semantically and syntactically meaningless but which preserved the normal phonological pattern of the language—sentences similar to Lewis Carroll's "Jabberwocky."[28] When these phonological strings were practiced, the maximum rate of speech was increased.

Study II. Practice at these three levels also facilitated the control of speech under DAF. Fewer errors in speech resulted when the *S*s practiced a sentence and then under DAF produced a translation than when they produced a nontranslation under DAF. The distinction between sentential and lexical meaning was also shown to be important for the control of speech under DAF: cross-language facilitation occurred for sentential meaning but not for lexical meaning.[29]

Study III. When a group of 21 *S*s under DAF produced completely meaningless sentences, they made more errors than in producing meaningful sentences, suggesting that meaning per se may facilitate the control of speech under DAF.

[28] L. Carroll, *Alice's Adventures under Ground: A Facsimile of the 1864 Manuscript*, 1965, 1–3.

[29] A theoretical interpretation of this finding appears in D. G. MacKay, Effects of ambiguity on stuttering: Towards a theory of speech production at the semantic level, *Kybernetik*, 5, 1969, 195–208.

AN EMPIRICAL SAMPLING STUDY OF THE DIXON AND MOOD STATISTICS FOR THE UP-AND-DOWN METHOD OF SENSITIVITY TESTING

By WILLIAM E. KAPPAUF, University of Illinois, Urbana

In the present study, empirical sampling procedures have been used to extend our information about the properties of the approximate maximum-likelihood estimators which Dixon and Mood developed for estimating μ, or $P_{.50}$, and σ of a probability-of-response function when data have been collected by the up-and-down method.[1] This method falls in the general class of the relatively new measurement methods which are variously described as sequential, adaptive, or staircase methods.[2] It is especially deserving of study and use by psychologists because it is well suited for determinations of the point of subjective equality (PSE) and because it will generally be more efficient in such applications than the traditional method of constant stimuli.

The up-and-down method is a routine for collecting observations. It applies in situations in which the subject is limited to two alternative responses, *e.g.* responses of "detection" vs. "non-detection," "second weight lighter than the first" vs. "second weight heavier than the first," etc. It is sequential in that the stimulus conditions

* Received for publication July 24, 1968. The work reported was supported in part by Research Grant NB-05576-01 from the National Institutes of Health. The present paper is a shortened version of a mimeographed report submitted to NIH: W. E. Kappauf, An empirical sampling study of the up-and-down method using computer-generated series, Report No. 1, April 4, 1967, 1–36. The author gratefully acknowledges the assistance of Carl E. Webber in computer programming and the participation of Frank R. Yeatman in the work of data summary and analysis. Thanks are due also to William Batchelder and Robert Bohrer for their helpful discussions of portions of this paper.

[1] W. J. Dixon and A. M. Mood, A method for obtaining and analyzing sensitivity data, *J. Amer. statist. Ass.*, 43, 1948, 109–126. There is also an earlier and more detailed document prepared by these authors but issued anonymously as a report of work under the Applied Mathematics Panel: Statistical analysis for a new procedure in sensitivity experiments, Appl. Math. Panel Report No. 101.1R, (Statistical Research Group, Princeton University, No. 40), July 1944, 1–58.

[2] T. W. Anderson, P. J. McCarthy, and J. W. Tukey, "Staircase" methods of sensitivity testing, NAVORD Report 65-46, Statistical Research Group, Princeton University, 1946, 1–134; M. M. Taylor and C. D. Creelman, PEST: Efficient estimates on probability functions, *J. acoust. Soc. Amer.*, 41, 1967, 782–787; G. B. Wetherill, Sequential estimation of quantal response curves, *J. roy. statist. Soc.*, 25, 1963, 1–48, Series B.

on each new trial are contingent upon the conditions and outcome of the previous trial. The first trial of an up-and-down series is conducted at a stimulus level chosen by the experimenter. If response X occurs on that trial, the second trial is run under stimulus conditions which, by one step on the stimulus scale, are more favorable to the occurrence of the opposed response, Y. If on the other hand, response Y occurs on the first trial, the next trial is run under stimulus conditions which, by one step on the stimulus scale, are more favorable to the occurrence of response X. All succeeding trials follow from the application of these same rules. Given that the probability of response X changes monotonically from 0 to 1.00 as successive steps are taken along the stimulus scale, these rules assure that trials will tend to concentrate in the vicinity of $P_{.50}$, where the probability of response X and the probability of response Y are equal. The general expectation is that relatively few trials will be conducted at extreme stimulus levels, because whenever the trial series moves to such a level, the response which is there highly probable drives the series back to less extreme levels. When a series of up-and-down observations has been terminated, the responses of each kind are tallied for each of the stimulus levels. Various computational procedures may be followed to arrive at estimates of $P_{.50}$ and the subject's variability.[3] The Dixon and Mood procedure is that of maximum likelihood. They considered the situation where the probability-of-response function may be accepted as a cumulative normal and where the stimulus levels used in testing are equally spaced on the scale for which normality applies. For this situation they developed estimators for μ, here identical with $P_{.50}$, and σ. More convenient to use by far are their approximate maximum-likelihood estimators, here designated m and s_{DM}. The maximum likelihood-estimators themselves, designated $\widetilde{m}$ and $\tilde{s}_{DM}$, require iterative evaluation.

While the up-and-down method has seen some use in psychological research,[4] it has not been used extensively. This probably is due

[3] Kappauf, Final Report on Context Effects in Psychophysical Judgments (Introduction and Part I, The Up-and-Down Method), Contract DA-49-007-MD-877, May, 1963, University of Illinois, 90; Wetherill, *op. cit.*, 21–24; Sharon Feeny, P. K. Kaiser, and J. P. Thomas, An analysis of data gathered by the staircase-method, this JOURNAL, 79, 1966, 652–654.

[4] See, for example, C. G. McDiarmid, Context effects in differential loudness judgments, unpublished doctoral dissertation, University of Illinois, 1962, 1–117; W. E. Kappauf, *loc. cit.*, Final Report, 1963; W. E. Kappauf, R. G. Burright, and W. DeMarco, Sucrose-quinine mixtures which are isohedonic for the rat, *J. comp. physiol. Psychol.*, 56, 1963, 138–143; P. T. Young and C. H. Madsen, Jr., Individual isohedons in sucrose-sodium chloride and sucrose-saccharin gustatory areas, *J. comp. physiol. Psychol.*, 56, 1963, 903–909.

to the fact that the method is not well known, but it may also be due in part to the fact that Dixon and Mood specifically caution their reader that their estimators strictly apply only when the number of observations in the up-and-down series is large, and that their approximate estimators should be used only if the step, d, between adjacent stimulus levels is no larger than 2σ. These restrictions leave unclear what one may expect if the approximate estimators are in fact used when series length is in the short and moderately-short range desired for most research. And they also leave unclear what measurement errors may be in store for an experimenter who wants to use large stimulus steps, either in pilot work or in testing, because he wishes to assure that the subject will have a reasonable number of judgments of low difficulty. The stated restrictions on the use of the approximate maximum-likelihood estimators arose because the original development was for asymptotic conditions, because m and s_{DM} were expected to be good approximations to $\tilde{m}$ and $\tilde{s}_{DM}$ only when d is less than 2σ, and because the standard errors of m and s_{DM} were expected to resemble those of $\tilde{m}$ and $\tilde{s}_{DM}$ only when d is less than 2σ. Dixon and Mood's specific comments were as follows:

> The method of statistical analysis of the data has been simplified considerably by the use of approximations which become poor for small samples. This puts a lower limit on the sample size if the present method of analysis is to be used. These considerations point to the need for using rather large samples. Samples of size 100 would do wery well for most purposes. If the sample size is less than 50, [this] method of analysis . . . should not be used.[5]

> A second condition on the experiment is that the sample size must be large if the analysis to be described is to be applicable. As it turns out, the effective sample size is only about half the actual sample size. The statistical analysis is based on large sample theory so that if one uses the analysis on a sample of size forty, he will in effect be using large sample theory on a sample of size twenty. Measures of reliability may well be very misleading if the sample size is less than forty or fifty.[6]

> The standard errors derived in this report apply to the maximum likelihood estimates and will not be applicable to [the approximate maximum-likelihood estimates] when they are different from [the maximum-likelihood ones].[7]

> It appears to be a formidable task to get the variances of the . . . [approximate maximum-likelihood] estimates, but for practical purposes the variances of the maximum likelihood estimates will serve when d is not much larger than [s_{DM}].[8]

[5] Dixon and Mood, *op. cit.*, Panel Report No. 101.1R, 1944, 10.
[6] Dixon and Mood, *op. cit.*, *J. Amer. statist. Ass.*, 43, 1948, 112.
[7] Dixon and Mood, *op. cit.*, Panel Report No. 101.1R, 1944, 48.
[8] Dixon and Mood, *op. cit.*, Panel Report No. 101.1R, 1944, 43.

The more complex analysis is required when the levels are not equally spaced or when the distance between levels exceeds twice the standard deviation. . . .[9]

Investigation of the properties of m and s_{DM} beyond the original work of Dixon and Mood has thus far been limited.[10] The papers by Brownlee, Hodges, and Rosenblatt and by Wetherill report some results on the problem of working with series shorter than 50 trials, but these data are limited in scope. It was thus in the context of obtaining systematic data over a wider range of measurement conditions that the present work was undertaken. In this study, empirical sampling distributions of m and s_{DM} were obtained by running extensive simulation trials on a computer. Three measurement conditions were varied in these trials: the step size, d; the number of trials per test; the phase relation of the set of test levels to μ. The data permit answers to the following questions which are important for considered applications of the up-and-down method in psychological research: How accurate and how reliable are m and s_{DM} as estimates of μ and σ if the number of observations is reduced to and below the working minimum of 40 or 50 trials suggested by Dixon and Mood? How do the estimates behave if d is extended beyond the recommended maximum size of 2σ? How satisfactorily can the variability of m and s_{DM} be estimated using the reliability formulae recommended by Dixon and Mood?

Dixon and Mood's estimators for μ and σ. When the responses recorded in up-and-down series are tallied by stimulus level, it becomes apparent that the tally of the X responses and the tally of the Y responses are redundant and indeed will at times be identical. They differ only to the extent that an up-and-down series happens to have started at a test level different from that which was reached as testing terminated. Because the experimenter's choice of initial level (whether near μ, or well above or below it) will influence the relative number of X and Y responses, Dixon and Mood adopted the procedure of estimating μ and σ from the data for the less frequent response. Their approximate maximum-likelihood estimators are given by the following formulae, wherein it is assumed that Y is the response which leads to an upward step in the up-and-down series

[9] Dixon and Mood, *op. cit.*, *J. Amer. statist. Ass.*, 43, 1948, 113.

[10] K. A. Brownlee, J. L. Hodges, and M. Rosenblatt, The up-and-down method with small samples. *J. Amer. statist. Ass.*, 48, 1953, 262–277; Kappauf, *loc. cit.*, Final Report, 1963; D. F. Votaw, The effect of non-normality on "staircase" methods of sensitivity testing (Statistical Research Group, Princeton University), May, 1948, 1–39; Wetherill, *loc. cit.*

and X is the response which leads to a downward step. N_{min} stands for the number of trials associated with the less frequent response. L stands for stimulus level, measured in stimulus units.

$$m\,|_{(N_Y = N_{min})} = (\text{mean test level at which } Y \text{ responses occurred}) + .5d$$
$$= \frac{\Sigma(f_Y)_i L_i}{N_{min}} + .5d$$

$$m\,|_{(N_X = N_{min})} = (\text{mean test level at which } X \text{ responses occurred}) - .5d$$
$$= \frac{\Sigma(f_X)_i L_i}{N_{min}} - .5d$$

$$s_{DM} = 1.620\, d(v + .029), \qquad [1]$$

where v is the variance in steps² (not in stimulus units²) of the test levels at which the less frequent response occurred, *e.g.* if $N_X = N_{min}$

$$v = \frac{N_{min}\Sigma(f_X)_i L_i^2 - (\Sigma(f_X)_i L_i)^2}{d^2 N_{min}^2}.$$

Dixon and Mood suggested that suitable approximations for the standard errors of m and s_{DM} should be obtained by using the formulae for the standard errors of the maximum-likelihood estimates themselves, $\tilde{m}$ and $\tilde{s}_{DM}$:

$$\sigma_{\tilde{m}} = \sigma G/\sqrt{N_{min}} \quad \text{and} \quad \sigma_{\tilde{s}_{DM}} = \sigma H/\sqrt{N_{min}}\,. \qquad [2]$$

Here G and H are 'look up' factors which express the dependence of the standard errors of the maximum-likelihood estimates, under asymptotic conditions, upon step size (d) and upon the phasing (ϕ) of the series of test levels with μ. Values of G and H for particular d and ϕ are read from published graphs.

If d/σ is not too large, m and s_{DM} approximate $\tilde{m}$ and $\tilde{s}_{DM}$ and hence one may assume that:

$$\sigma_m = \sigma G/\sqrt{N_{min}} \quad \text{and} \quad \sigma_{s_{DM}} = \sigma H/\sqrt{N_{min}}\,. \qquad [3]$$

In practice, of course, σ is not known and the estimation formulae become:

$$SE_m = s_{DM} G/\sqrt{N_{min}} \quad \text{and} \quad SE_{s_{DM}} = s_{DM} H/\sqrt{N_{min}}\,. \qquad [4]$$

Method

The analyses which follow concern: the three measurement parameters N_{min}, step size (d), and phase (ϕ); the statistics m and

s_{DM}, and their empirical average values $\bar{m}$ and $\bar{s}_{DM}$; the empirical standard errors of m and s_{DM}, represented as SD_m and $SD_{s_{DM}}$; the factors G and H, and their empirical counterparts here designated G' and H'. Step size is defined by the relation: $d = L_{i+1} - L_i$. Phase condition is given numerically as: $\phi = [(\text{test level nearest } \mu) - \mu]/d$.

Computer generation of up-and-down series. Up-and-down observations under the normal model can be simulated by referring to a table of random normal deviates. Take any starting level in normal-deviate form and any step size in σ units. To conduct a 'trial,' enter the table at some random point and observe whether the random deviate listed there is algebraically larger or smaller than the starting level. If larger, the 'response' or outcome of that 'trial' is one which demands an upward step, which means that the next simulated trial should be conducted at the starting level plus one step. If smaller, the response or outcome is one which demands a downward step, meaning that the next simulated trial should be conducted at the starting level minus one step. Conduct the new trial in the same manner as the first, *i.e.* observe whether the next deviate listed in the table is larger or smaller than the new test level. This procedure of consulting printed tables to generate up-and-down series was used in empirical sampling studies by Votaw and by Guth and Kappauf.[11] In both cases, however, slowness of the operation imposed limits on the amount of data which could reasonably be collected.

The present study, like Wetherill's, takes advantage of the fact that a modern computer makes it possible to generate and analyze up-and-down series with great speed.[12] The computer used was an IBM 7094 computer in the University of Illinois Computer Science Laboratory. It was programmed to generate a random normal deviate, compare it with the current test level, move the series up or down appropriately, continue until a designated N_{min} had been reached, and then to summarize the series. It calculated m for the series (and stored it in four digits as ±X.XXX), calculated v (stored as .XXXX), and noted the number of stimulus levels which had been used (stored as X). Generation of the next series began immediately, starting at the test level reached at the end of the previous series. The program required that 200 series be completed under any given set of conditions of phase, step size, and series length before one or more of these test parameters was changed and a new block of series was begun. There were in all 176 experimental conditions, as defined by the following levels of the program parameters: $\sigma = (1.00)$; $\mu = (0.00)$; $N_{min} = (10, 20, 40, 80)$; $d = (.50, .75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50, 3.00, 4.00, 6.00)$; $\phi = (-.25, .00, +.25, +.50)$.

Data output. Summary statistics for each up-and-down series were recorded on tape and transferred to punched cards. Subsequently the cards were reintroduced in the computer and the data for each block of 200 series were compiled to obtain the empirical sampling distributions of m and s_{DM} for that condition. The mean and standard deviation of each of these empirical sampling distributions were computed and tabulated, thus providing $\bar{m}$, SD_m, $\bar{s}_{DM}$, and $SD_{s_{DM}}$, for each of the 176 conditions under study.

[11] Votaw, *op. cit.*, 13–15; see Kappauf, *op. cit.*, Final Report, 1963, 28.
[12] Wetherill, *loc. cit.*

Results

Some Anticipated Relationships

Some results for m, s_{DM}, and their distributions can be inferred at the outset. It is useful to review these in advance of the empirical data.

Discrete distributions of m and s_{DM}. Because trials in an up-and-down series occur at specific levels which are d apart, the distributions of m, v, and s_{DM} are discrete. For any given set of measurement conditions, successive possible values of m differ by d/N_{min}. Successive possible values of s_{DM}, given m, differ by $3.24d/N_{min}$. For d small and N_{min} large, s_{DM} and m may each take on many values. For d large and N_{min} small, they can take on but few values.

Limiting relations when d is inappropriately large. (1) If $\phi = .00$ and d is very large, half of the trials will be conducted at μ and the remainder will be binomially distributed at the levels just above and just below μ. m must be unbiased, while s_{DM} must grow linearly with d, as must both SD_m and $SD_{s_{DM}}$. (2) If $\phi = .50$ and d is very large, successive trials will always straddle μ. v will always be zero regardless of N_{min}, and so the value of s_{DM} will be a function of the constant term in the equation for s_{DM}: it will, in fact, be equal to $.047d$. The value of m will always be identically equal to μ. Clearly SD_m and $SD_{s_{DM}}$ will both be zero. Thus these quantities must approach zero in the present empirical runs as d grows larger. (3) If $\phi \neq .00$ or. 50 and d is very large, successive trials will again straddle μ, but in this case m will be biased. The other three statistics will behave as they did for the case where $\phi = .50$; that is, s_{DM} will approach $.047d$ and both SD_m and $SD_{s_{DM}}$ will approach zero.

Bias in m. For any combination of d and ϕ it is possible to find the asymptotic or stationary distribution of trials under the normal model. Computing m for these distributions, one finds that m is always biased when ϕ is other than .00 or .50. The bias is small when $d < 2\sigma$ but grows for $d > 2\sigma$. Thus for $\phi = +.25$ and $\sigma = 1$, $(m - \mu)$ is $-.00001$ for $d = 2\sigma$, $-.105$ for $d = 3\sigma$, $-.371$ for $d = 4\sigma$ and -1.10 for $d = 6\sigma$.

Empirical Properties of s_{DM}[13]

Bias. Values of $\bar{s}_{DM}$ are shown in Fig. 1 for $\phi = .00$, .50, and .25. In the latter case, values of s_{DM} for $\phi = +.25$ and $-.25$ have been

[13] A complete tabulation of the data on which this paper is based appears in the mimeographed report cited in footnote 1 above.

averaged. Connected open circles represent data for $N_{min} = 80$. Connected solid circles are for $N_{min} = 10$. Inasmuch as σ was 1.00 throughout, it is seen that: (1) When $d < 2\sigma$, s_{DM} typically underestimates σ. This bias is present for all phase conditions. It is small when N_{min} is 80, increases as N_{min} becomes smaller, and can be of the order of 20% when $N_{min} = 10$. (2) Over the range where $d < 1.75\sigma$, $\bar{s}_{DM}$ grows with d. The constant term 0.029 in the formula for s_{DM} contributes $.047d$ to the value of s_{DM}, and as Fig. 1 shows, for small d, $\bar{s}_{DM}$ grows with d in a manner which roughly parallels this contribution. These results and those in (1) immediately above are consistent with Wetherill's finding of negative bias in the estimation of σ for cases with $N_{min} = 17$, $\phi = 0$, and $d = .5, 1.0, 2.0$.[14] This bias decreased with increases in d. (3) When $d > 2\sigma$, s_{DM} is positively biased when $\phi = .00$, negatively biased when $\phi = .50$, and relatively unbiased when $\phi = .25$. Dixon and Mood were aware of this dependence of s_{DM} on ϕ, and they even suggested that s_{DM} might be corrected on the basis of phase when v is small.[15] The expected value of v is small only when d is large, so this proposed correction functions as an adjustment for having taken d too large. Clearly, however, such a correction routine is unrealistic when N_{min} is small or when d is very large, because under these conditions phase information is itself not sufficiently dependable to serve as the basis for a good adjustment.

The standard error of s_{DM}. Because the standard error of s_{DM} depends, as all standard errors do, upon $\sqrt{N}$, it is more informative here to examine the behavior of $\sqrt{N_{min}}(SD_{s_{DM}})$ than to consider $SD_{s_{DM}}$ itself. This quantity, $\sqrt{N_{min}}(SD_{s_{DM}})$, corresponds to Dixon and Mood's H. If, from Equation 3, $\sigma_{s_{DM}} = \sigma H/\sqrt{N_{min}}$, then for the present case where $\sigma = 1.00$, it is reasonable to define H' by: $H' = \sqrt{N_{min}}(SD_{s_{DM}})$.

Fig. 2 presents the obtained values of H' as a function of d, ϕ, and N_{min}. Data for $\phi = +.25$ and $\phi = -.25$ have again been averaged. Smoothed functions were fitted by eye to represent the data for $N_{min} = 10$ and 80. These functions may be compared with those which Dixon and Mood obtained for H, given N_{min} large (see the dashed lines in Fig. 2). From Fig. 2, it may be seen that: (1) H' varies with d in a manner which depends critically on ϕ, as had been known for H. (2) H' depends on N_{min}, the general rule being that H' is smaller when N_{min} is small. This is in whole or in part a secondary effect, however, for it is associated with the fact that s_{DM}

[14] Wetherill, *op. cit.*, 16.
[15] Dixon and Mood, *op. cit.*, Panel Report No. 101.1R, 1944, 16.

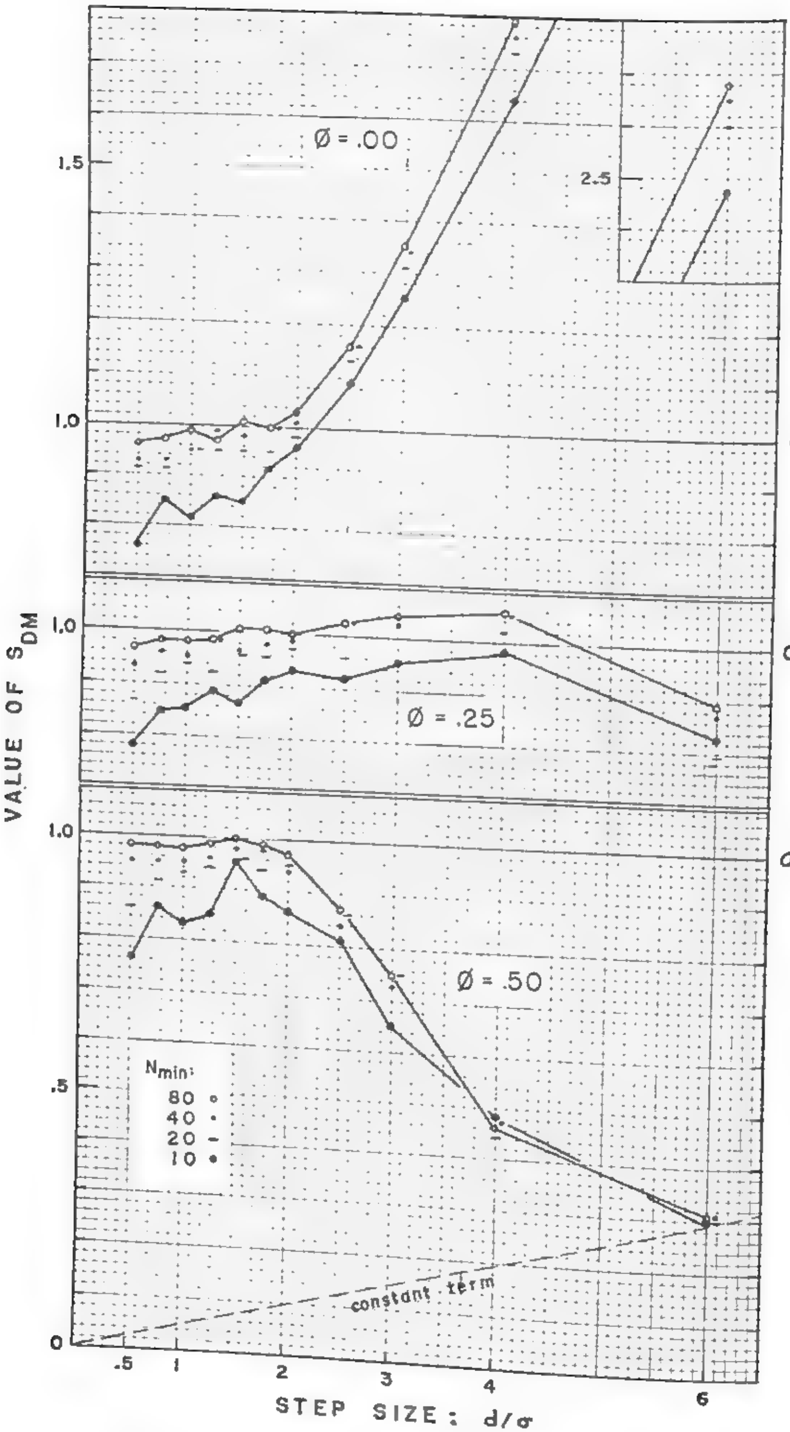


FIG. 1. VALUES OF $\bar{s}_{DM}$ AS A FUNCTION OF d, ϕ, AND N_{min}

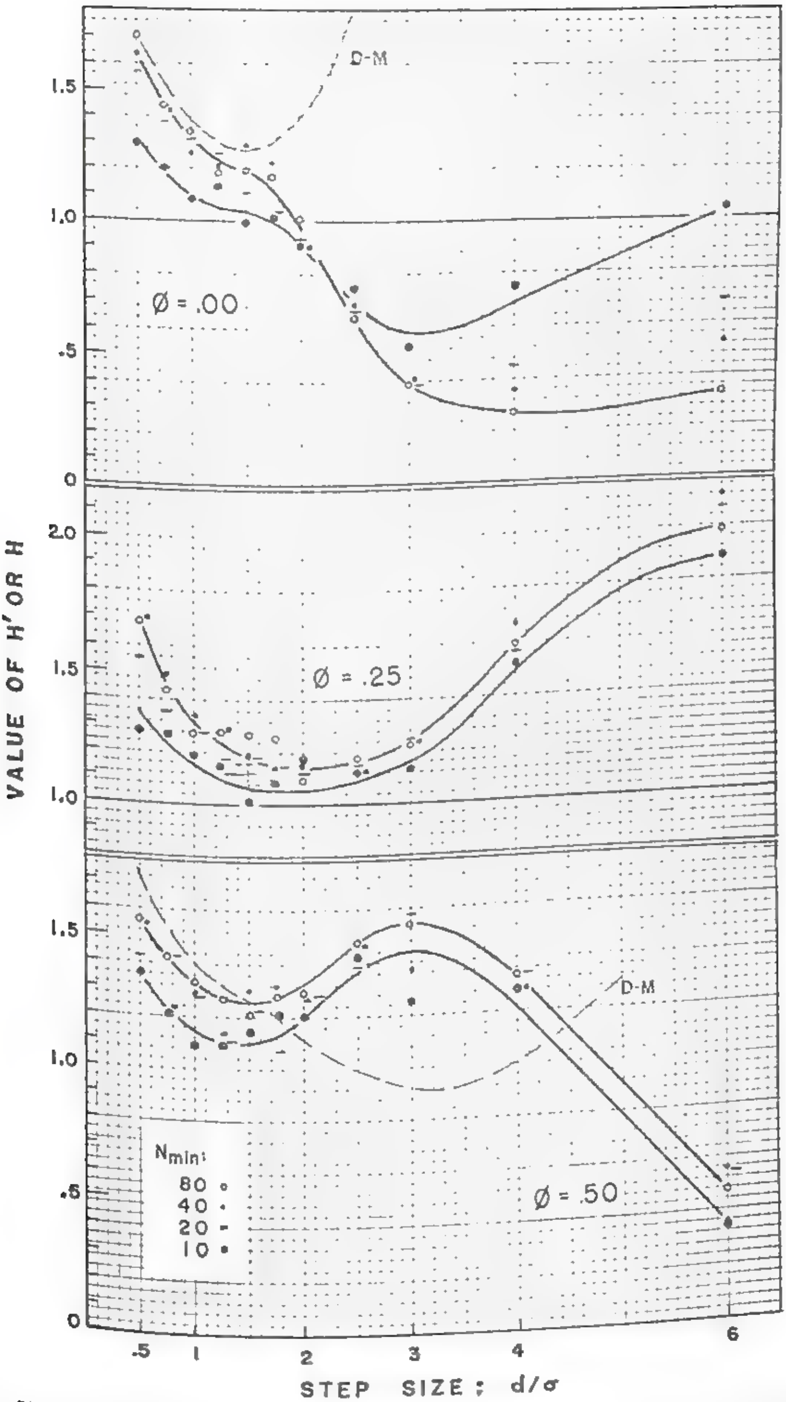


FIG. 2. VALUES OF H', *i.e.* $\sqrt{N_{min}}\,(SD_{s\,p\,M})$, AND VALUES OF H AS A FUNCTION OF d, ϕ, AND N_{min}

(Dashed lines are H functions from Dixon and Mood.)

is typically smaller when N_{min} is small (see Fig. 1). (3) When $d < 1.5\sigma$, values of H are larger than but very similar to the companion values of H' for $N_{min} = 80$. Thus, even though the H functions were derived for the maximum-likelihood estimates of σ and not for s_{DM}, they provide a reasonably sound basis for estimating the standard error of s_{DM} when N_{min} is large and d is small, as Dixon and Mood had forecast. (4) When $d > 1.5\sigma$, there is conspicuous dissimilarity between the functions for H derived by Dixon and Mood and the functions for H' found here. Two factors appear to be involved in this difference. First, the Dixon and Mood values apply to the actual maximum-likelihood estimates, while the present data apply to the approximation estimates. Second, the derivation of H, and of G also, presumed a negligible correlation between m and s_{DM}, a condition which does prevail for $d < 2\sigma$. Evidence from the empirical data for m and s_{DM}, however, would indicate that the correlation becomes significantly large as d grows larger than 2.5σ.

Empirical Properties of m.

Bias. Empirical values of $\bar{m}$ are shown in Table I. Inasmuch as μ was .00 in all cases, m itself tells the bias story. These data agree closely with the expected bias values cited above for $\phi = .25$. For all practical purposes, m is unbiased regardless of phase when $d \leq 2\sigma$.

The standard error of m. It is appropriate here to employ the same strategy as that used above to consider the standard error of s_{DM}, *i.e.* to eliminate the effect of N_{min} on the standard error of m by defining G' as $\sqrt{N_{min}}(SD_m)$. It is interesting to note that the standard error of G' is independent of N_{min} and for the present empirical data is approximately $.05G'$:

$$SE_{G'} = \sqrt{N_{min}}\,(SE_{SD_m})$$
$$\approx \sqrt{N_{min}}\left(\frac{SD_m}{\sqrt{2(200)}}\right)$$
$$\approx .05\,G'.$$

This error figure has been used to test the homogeneity of the G' values obtained for the four N_{min} conditions under each (d, ϕ) condition. Of the 44 sets of G' values tested, the number 'rejected' as not homogeneous by a .05-level range test was only five, which itself is a nonsignificant (.05-level test) number of rejections. All sets were therefore assumed homogeneous, and each set of four G' values was averaged across N_{min} conditions. These averages are the values

TABLE I

VALUES OF $\bar{m}$ OBTAINED FROM COMPUTER-GENERATED SERIES WITH $N_{min} = 80$, TABLED AS A FUNCTION OF STEP SIZE AND PHASE

ϕ	Step size: d/σ										
	.50	.75	1.00	1.25	1.50	1.75	2.00	2.50	3.00	4.00	6.00
−.25	.004	.007	.013	−.000	.001	−.005	.011	.029	.099	.365	1.083
.00	.015	.021	.015	.017	−.017	−.014	.008	.010	.003	.003	.017
.50	.024	.001	−.013	.008	.004	.015	.009	.003	.000	−.005	−.001
+.25	−.010	−.002	−.004	.008	−.002	.000	−.002	−.032	−.106	−.370	−1.102

Note: $\mu = .00$, and so deviation from .00 is the amount of bias. Scale can be appreciated from the fact that $\sigma = 1.00$. 200 separate empirical values of m are represented by each tabled value of $\bar{m}$.

plotted in Fig. 3. The data for the four ϕ conditions are plotted separately.

Fig. 3 therefore shows the dependence of the standard error of m upon d and ϕ after the effect of N_{min} has been removed. Smoothed functions were fitted by eye to the empirical data. The dashed lines represent the Dixon and Mood G functions. It will be seen that: (1) For $d < 2\sigma$, the values of G and G' agree very well, and G' is independent of ϕ, as Dixon and Mood advised it should be. (2) For $d > 2\sigma$, G' varies markedly with ϕ. When $\phi = .50$, G' declines as d becomes so large that successive trials typically straddle μ. For $\phi = .00$, the G' function coincides with the line $G' = .5d$ when d is large, because for this phase condition the variability of m depends

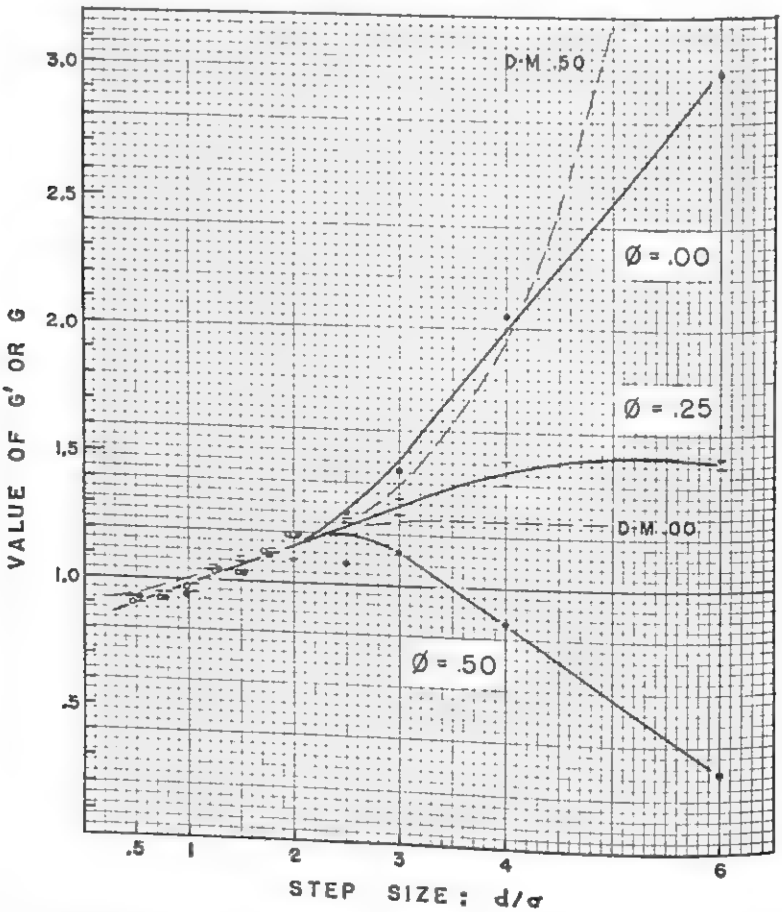


Fig. 3. Values of G', *i.e.* $\sqrt{N_{min}}\,(SD_m)$, and Values of G as a function of d and ϕ

(See text for averaging across levels of N_{min}.
Dashed lines are G functions from Dixon and Mood.)

purely on the standard error of P (where $P = Q = .50$). (3) For $d > 2$ or $2.5\ \sigma$, there is disagreement between the functions for G and G' for the different phase conditions, just as there was for H and H' and apparently for the same reasons. (4) For d between $.5\sigma$ and 2σ, the dependence of G' on d is very closely approximated by a straight line.

Discussion

It is clear from the data here presented that when series length is short and N_{min} small, s_{DM} can never provide a useful estimate of σ. While the interest in estimating σ for a probability-of-response function is never so great as the interest in estimating μ or $P_{.50}$, there are nevertheless two critical and important uses for a good estimate of σ. One is in the selection of an appropriate stimulus step. The other is in estimating the standard error of m. The data say that one can never expect to select step size wisely on the basis of a short, pilot up-and-down series. If the step adopted for these pilot observations happens to be small, s_{DM} will be negatively biased as well as unreliable. And if the step happens to be large, the value of s_{DM} may be seriously biased in either a positive or a negative direction depending upon the prevailing phase condition. Choice of step size therefore requires extensive background testing or alternative ways of assessing d than simply in terms of s_{DM}.

There is one previous study of the adequacy of G for the prediction of the standard error of m. This is the study of Brownlee, Hodges, and Rosenblatt, in which the variability of m was investigated for very small samples and for d relatively small, *i.e.* $.67\sigma$, 1.0σ and 1.5σ.[16] It was found that *when σ is known* the Dixon and Mood formula overestimates the variability of m slightly, implying G to be conservatively large. The amount of overestimation is somewhat greater the smaller d is. This result is in agreement with the trend in Fig. 3, where G exceeds G' by a small and seemingly increasing amount as d becomes smaller.

In a psychological experiment concerned with the determination of a threshold or PSE, considerable interest usually attaches to SE_m. Suppose G is taken to be a linear function of d/σ, for d/σ less than 2.0. Then a function which appears reasonable from the empirical data in Fig. 3 is: $G' = .82 + .16d/\sigma$, and from Equation 4 SE_m becomes:

$$SE_m = s_{DM}(.82 + .16d/s_{DM})/\sqrt{N_{min}}$$
$$= (.82s_{DM} + .16d)/\sqrt{N_{min}}.$$

[16] Brownlee, Hodges, and Rosenblatt, *loc. cit.*

In this form, the formula makes explicit the relative contributions of s_{DM} and d to SE_m. It also makes explicit the fact that bias in s_{DM} will necessarily introduce bias in the estimate of SE_m. Thus, for cases where the step size is in the range between 1σ and 2σ as recommended by Dixon and Mood and the number of observations is close to the advised minimum of 50, SE_m will be underestimated due to the negative bias of s_{DM}.

Some Clues to Improved Estimates of μ and σ. The data here presented also suggest that several possibilities exist for the development of easy-to-calculate estimates of μ and σ which would have properties more desirable than those of m and s_{DM}. (1) It is clear that the bias in s_{DM} when N_{min} is small and when d is less than 2σ is of the sort which might be expected to yield to a multiplier of the form $N/(N-a)$. Inasmuch as the multiplier $N/(N-1)$ transforms a sample variance into an unbiased estimate of the population variance, a similar multiplier might have a comparably desirable effect in the present up-and-down situation. (2) It is noteworthy that the bias in s_{DM} when d is large develops in opposite directions under what may be termed complementary phase conditions, *i.e.* for $\phi = .00$ and $\phi = .50$, while the bias in m for d large also develops in opposite directions under complementary phase conditions, in this case for $\phi = +.25$ and $\phi = -.25$. These observations suggest that relatively unbiased estimates of both μ and σ might be obtained if one were to (a) take observations in two concurrent up-and-down series with $d_1 = d_2$ but with stimulus levels differing so that $\phi_1 = \phi_2 + .50$, (b) estimate μ by averaging the values of m from these two complementary series, and (c) estimate σ by using a similar composite of measures from the two series. These leads are being examined and a report on the results of that inquiry will follow shortly.

Summary

Computer simulation runs have been used to provide information about the properties of the approximate maximum-likelihood estimators of μ and σ which Dixon and Mood developed for use with the up-and-down method. Certain of the results bear out the conclusions and expectations expressed by Dixon and Mood. Other results provide information which could not be obtained except through empirical sampling procedures.

Outcomes which are directly related to and confirm the Dixon and Mood analysis. When stimulus step size, d, is greater than 2σ,

both m and s_{DM} are subject to serious bias when particular phase conditions prevail. Their standard errors are uncertain, for they too depend on phase condition. m and s_{DM} are not, therefore, useful estimates of μ and σ when d is large; the Dixon and Mood recommendations are substantiated. When step size is smaller than 2σ, m is unbiased or effectively so for all conditions of phase and N_{min}. The standard error of m increases gradually with d in a manner which is independent of phase. The Dixon and Mood values for G used for estimating the standard error of m are very satisfactory, though conservatively large by a small factor. s_{DM} and its standard error are sensitive to reductions in N_{min} as Dixon and Mood presumed they might be but could not quantify.

Outcomes which provide new information. When step size is smaller than 2σ, s_{DM} is essentially unbiased if N_{min} is as large as 80, but it becomes increasingly biased in the negative direction as N_{min} grows smaller. This negative bias is independent of phase but depends on d. The latter effect is associated primarily with the second constant in the formula for s_{DM}, which contributes to s_{DM} an amount which is a multiple of d. When step size is less than 1.5σ, the standard error of s_{DM} can be estimated well from s_{DM}, in spite of its negative bias, if H which is consistently larger than H' is used in the prediction formula. When the step size is between $.5\sigma$ and 2σ, it should be adequate for most purposes to estimate the standard error of m using a value of G which is a simple linear function of d. Estimates of the standard error of m will typically underestimate the standard error of m because of the negative bias in s_{DM}. For large step sizes, where Dixon and Mood cautioned against using m and s_{DM}, the standard errors of these estimates are not well described by the published functions for G and H.

VISUAL LIGHTNESS ASSIMILATION AND CONTRAST AS A FUNCTION OF DIFFERENTIAL STIMULATION

By Joseph A. Steger, State University of New York at Albany

The phenomenon of simultaneous lightness contrast[1] has been the concern of many experimenters.[2] Its reversal, herein referred to as assimilation, has not been of such immediate experimental concern, even though it was demonstrated as early as 1876.[3] The same disproportion is found in the theoretical area, for although many contemporary authors have addressed themselves to theoretical explanations of contrast,[4] few have concerned themselves with theoretical accounts of assimilation.[5] Nevertheless, following his earlier work on assimilation, Helson developed an hypothesis, based on the differences of stimulation in neighboring retinal areas, that synthe-

* Received for publication October 19, 1968. The paper is based on a portion of a doctoral dissertation done at the Psychology Department, Kansas State University. Particular appreciation is extended to Harry Helson for his personal guidance and criticism throughout the research.

[1] Whenever contrast and assimilation are referred to in this paper, it can be assumed that only simultaneous lightness contrast and assimilation are meant. Any other type of contrast or assimilation is specified

[2] M. Alpern, Simultaneous brightness contrast for flashes of light of different durations, *Invest. Ophthalmol.*, 2, 1963, 47–54; P. W. Berman and H. W. Liebowitz, Some effects of contour on simultaneous brightness contrast, *J. exp. Psychol.*, 69, 1965, 251–256; O. Bryngdahl, Perceived contrast variation with eccentricity of spatial sine-wave stimuli, *Vision Res.*, 6, 1966, 553–565; A. L. Diamond, Foveal simultaneous brightness contrast as a function of inducing- and test-field luminances, *J. exp. Psychol.*, 45, 1953, 304–314; E. H. Heineman, The relation of apparent brightness to the threshold for differences in luminance, *J. exp. Psychol.*, 61, 1961, 389–399; H. W. Horeman, Inductive brightness depression as influenced by configurational conditions, *Vision Res.*, 3, 1963, 121–130; D. Jameson and L. M. Hurvich, Complexities of perceived brightness *Science*, 133, 1961, 174–179; and J. Nachimas, Effect of exposure duration on visual contrast sensitivity with square-wave gratings, *J. opt. Soc. Amer.*, 57, 1967, 421–427.

[3] W. Von Bezold, *The Theory of Color and its Relation to Art and Art-Industry* (S. R. Koehler, trans.), 1876. For a review of the literature dealing with assimilation, see J. Steger, The reversal of simultaneous contrast, *Psychol. Bull.*, 70, 1968, 774–781.

[4] A. L. Diamond, A theory of depression and enhancement in the brightness response, *Psychol. Rev.*, 67, 1960 168–199; G. A. Fry, Mechanisms subserving simultaneous brightness contrast, *Amer. J. opt. Arch. of Amer. Acad. Opt.*, Whole No. 45, 1948; D. Jameson and L. M. Hurvich, Theory of brightness and color contrast in human vision, *Vision Res.*, 4, 1964, 135–154; and H. K. W. Moed, Constancy and contrast: Parts I, II, and IIIa, *Acta Psychol.*, 22, 1964, 272–320; 24, 1965, 91–166; 25, 1966, 222–292.

[5] See Steger, *loc. cit.*

sizes both contrast and assimilation.[6] According to this hypothesis: "We therefore assume that relatively equal impulses from neighboring areas summate, or at least do not inhibit one another, thereby giving rise to assimilation, while an intense excitation, arriving at some common synapse or neural pathway inhibits a weaker one to give rise to contrast."[7] As corollary to this, Helson added the area-luminance hypothesis, according to which area acts like intensity, *i.e.* an increase in area has the same effect as an increase in luminance. The full development of this theoretical model is stated in terms of differential stimulation: low differentials of stimulation summate, yielding assimilation; and high differentials of stimulation inhibit, yielding contrast. Helson also pointed out that there should be an intermediate differential of stimulation, in which neither summation nor inhibition occurs and, hence, neither contrast nor assimilation is found. This deduction was verified experimentally by Helson and Joy.[8] Verification of the stimulation-differential model has been demonstrated for three variables—line and space widths[9] and background reflectance[10]—but these studies have employed only white or black lines.

A more direct test of the concept of stimulation differentials was made in the present study, by use of intermediate values of line and background reflectances coupled with variation of line and space widths. The present study also provided a more direct test of the area-luminance hypothesis, specifically of its predictions that small differentials in stimulation will yield assimilation; large differentials, contrast; and intermediate differentials, neither assimilation nor contrast. Thus, the present study varied all the stimulus dimensions (line reflectance, line width, line separation, and background reflectance) in an effort to describe the response continuum as a function of the stimulus variables and their interactions in terms of Helson's differential-stimulation model.

METHOD

To assess the independent variables quantitatively, the same experimental technique employed in earlier studies was used.[11] Overlaid on cards of speci-

[6] H. Helson, Studies of anomalous contrast and assimilation, *J. opt. Soc. Amer.*, 53, 1963, 179–184; Helson and V. L. Joy, Domains of lightness assimilation and contrast, *Psychol. Beitr.*, 6, 1962, 405–415 (Engl. trans.); and Helson and F. G. Rohles, Jr., A quantitative study of reversal of classical lightness-contrast, this JOURNAL, 72, 1959, 530–538.

[7] Helson, *op. cit.*, 183.

[8] Helson and Joy, *loc. cit.*

[9] Helson and Joy, *loc. cit.*

[10] Helson and Rohles, *loc. cit.*; Helson and Joy, *loc. cit.*

[11] Helson, *loc. cit.*

fied background reflectance were vertical lines of a given width, separation, and reflectance. On each side of the card were lines of different reflectances, and these halves were judged by the method of paired comparisons in terms of a category scale of lightness. Thus one half of the card was the standard and the other the variable; these were reversed on alternate presentations. In the present study, three background reflectances and four line reflectances were selected to sample the reflectance continuum. Three values of line and space widths were also selected on the basis of previous findings that large line and space widths (above 15 mm.) yield contrast, small (below 5 mm.) yield assimilation, and intermediate (10 mm.) yield neither assimilation nor contrast.

Since all of these independent variables act simultaneously in any given stimulus complex, the interactions were also important. The use of two slightly different experimental techniques aided assessment of the interactions of line and space widths and line and background reflectances. With Technique A, one half of each card was overlaid with one line reflectance and the other with a different line reflectance, but both reflectances were of equal line and space widths. This provided a comparison of the effects of varying line reflectance, while holding all other independent variables equal. Since this was also the method used in earlier studies, it allowed comparison with their findings. Technique B varied from A in that only one half of the card was ruled with lines; the other half was a homogeneous gray. Comparisons with this technique were thus not complicated by two different sets of independent variables. This type of presentation also provided one gray standard throughout the session for comparison of the independent-variable combinations.

Subjects. Ss were 15 undergraduate and 5 graduate students enrolled at Kansas State University. None of these Ss had had previous experience with judgments of the type required in the present study. They were 18–24 years old, and all had normal vision or wore corrective lens. Ten Ss were used with each procedure.

Stimulus materials and apparatus: Technique A. The stimulus materials consisted of 162 cards, 7 × 11 in.; 54 of these cards having a reflectance of 65%, 54 a reflectance of 36%, and 54 a reflectance of 9%. These cards were cut from Poster Board No. 51 and Crescent Board Nos. 651 and 864, respectively. On each of the 162 cards were overlaid vertical lines such that each card had two halves with lines of different reflectance but of equal line and space widths. The reflectances of the lines were either 83%, 37%, 20%, or 4%. These overlaid lines were cut from Color Vue papers Nos. 1, 7, 11, and 18, respectively, into widths of 3, 9, and 16 mm. Each of these line widths was glued to the cards with space widths of 3, 9, and 16 mm. Each line reflectance was used with each line and space width to overlay one half of the card while a different line reflectance of the same line and space width overlaid the other half. This made it possible to compare the two card halves simultaneously. Thus the following line-reflectance combinations were employed: 83%–37%, 83%–20%, 83%–4%, 37%–20%, 37%–4%, 20%–4%.

A three-sided booth was constructed, each side 48 in. wide × 60 in. high. Each of the three sides was lined with a homogeneous lightweight cardboard matching one of the background reflectances. This was accomplished by covering a wooden frame with the same cardboard as that used for the cards.

Cards of the same reflectance as the surround were presented with the long axis horizontal in an aperture $10\frac{3}{4} \times 6\frac{3}{4}$ in. having its center 24 in. from the edge of the booth and 40 in. from the floor. A slide was so constructed behind the aperture that a stimulus card could be presented for a period of 2 sec. When removed, it was automatically replaced by a plain card of the same reflectance. The booth provided a quick means of randomly varying the backgrounds and presentation of the stimuli as well as serving to shield the stimuli and *E* while he recorded *S*'s responses. The windowless experimental room was evenly illuminated by two banks of Cool White 40-w. fluorescent lamps (approximately 100 footcandles), thus assuring uniform achromatic illumination at all times.

—*Technique B.* The stimuli were the same as those in A with the exception that only one half of the card was presented, the other half being covered by a plain surface of the same reflectance as that of the intervening gray of the lined side. The booth and experimental room were the same as those in Technique A.

Procedure: Technique A. This was developed from a modification of the complete method of paired comparisons. Only lines of the same width and separation on the same background were compared, rather than every line and space width and background with every other as in the usual method of paired comparison. The usual equation for the number of comparisons in the method of paired comparisons is $n(n-1)/2$, but in the present study this equation held only in the case of line reflectances since we had to include equal line and space widths. Also included were the reverse combinations of line and space widths, since a given line and space combination is not equivalent when the line assumes the space dimension and the space the line dimension. For example, a line width of 3 mm. and space width of 9 mm. is not the same stimulus combination as a line width of 9 mm. and a space width of 3 mm. But in the case of line reflectances a combination of 83% and 37% is equivalent to a combination of 37% and 83%. The total number of pairs was therefore equal to 3 backgrounds × 6 line-reflectance combinations × 3 line widths × 3 space widths, or 162. Although the previous studies used the method of limits in presenting the stimuli to aid in reducing response variability,[12] the stimuli—line reflectances, widths, and space—for each background condition in the present experiment were presented in random order based on a table of random numbers. This was done to avoid order effects and meet the assumptions required in the statistical analysis. The background order was also randomized for each *S*.

The *S* was seated at a viewing distance of 3 m. and given a card that was of the same reflectance as the stimuli to be viewed and on which were written these categories: Very, very much lighter; Very much lighter; Much lighter; Lighter; Equal; Darker; Much darker; Very much darker; and Very, very much darker. These instructions were then given:

> The purpose of this experiment is to study the way in which certain grays are perceived. I will show you a series of cards like the one I have now. [*E* displayed a card not used in the present study which yielded 100% agreement in preliminary observations.] Tell me, how does the right side look

[12] Helson, *loc. cit.*

> compared to the side on your left using the categories on the card I just gave you?
>
> [After S responded, E continued:] This is exactly what you are to do as I show you the cards in the window in front of you. If I ask you to judge to your right you report how the gray on your right appears with respect to the gray on your left.
>
> Feel free to use all or any of the categories or to ask questions should you have any during the course of the experiment.

The first *S* asked what he was to compare when cards of large line widths (16 mm.) and small space widths (3 mm.) appeared. This was eliminated as a possible source of confusion by telling this *S* and all other *Ss* which areas were to be compared. This was done during the instruction period in the case of the other *Ss* by providing such a card and noting the areas to be compared, thus avoiding difficulty as to what was line and what was background. Since it took approximately an hour to repeat the 51 comparisons four times within each background, Technique A was divided into three sessions of one hour each, approximately a day apart.

—*Technique B*. This was a modification of that used in A. Since in B only half the card was overlaid with vertical lines, the nine combinations of line and space widths within each of the four line reflectances yielded 36 card halves for each of the three backgrounds, making a total of 108 comparisons (3 line widths $\times$ 3 space widths $\times$ 4 line reflectances $\times$ 3 backgrounds). *Ss* were given the same category scale as that used with Technique A.

The modified procedure of Technique B is best understood from the instructions given *S*:

> The purpose of this experiment is to study the way in which certain grays are perceived. I will show a series of cards like the one present in the opening directly in front of you.
>
> I would like you to tell me how the gray with the lines or stripes on it appears in relation to the plain gray on the other side.
>
> It is obvious I don't want you to tell how the lines look but how the gray between the lines appears. Consider the plain gray side as the reference and judge how the line gray appears in relation to the plain gray using the categories provided.
>
> Do you have any questions?

Since the plain side in Technique B served as the standard, its position was counterbalanced by having each *S* judge each card twice on his left and twice on his right with the left or right starting order split evenly between the 10 *Ss* and assigned by chance. The only other modification was in presentation order of the stimuli. Rather than random presentation as in A, the cards were shown in an ascending and descending order by line width and space width from largest or smallest lines and spaces.

Treatment of the data.[13] The same experimental design, viz. a modified split-plot, was employed with both techniques. The backgrounds served as plots split into line-reflectance combinations in which selected line and space widths were present. This design allowed a statistical analysis of the main effects and, more importantly, of the interactions of all factors. For both techniques, *Ss'* category-scale responses were transformed into numerical values in the following manner: a response of 'Equal' was given a score of 5.

[13] The statistical guidance of H. Feyerherm is gratefully acknowledged.

responses indicating contrast were given scores below 5 (4 = Darker or Lighter, 3 = Much darker or lighter, 2 = Very much darker or lighter, 1 = Very, very much darker or lighter, and 0 = the most extreme category if *Ss* wished to use it), and responses indicating assimilation were given scores above 5 (6 = Darker or Lighter, 7 = Much darker or lighter, etc.). Judgments of replicated stimulus conditions were averaged for each *S*. These averages were then summed and averaged for the group. The mean judgment for each stimulus was then available. Each mean therefore represents 40 observations with Technique A and 40 with Technique B. Another mean for each line reflectance within each background was computed combining line widths and separations. An analysis of variance was made of the data found with Techniques A and B to evaluate the results.

Results

Technique A. The results of Technique A clearly verify the differential-stimulation hypothesis as well as the notion that responses of assimilation and contrast are complementary.

Fig. 1 and Table I show that three different response functions were obtained, one for each of the three background reflectances.

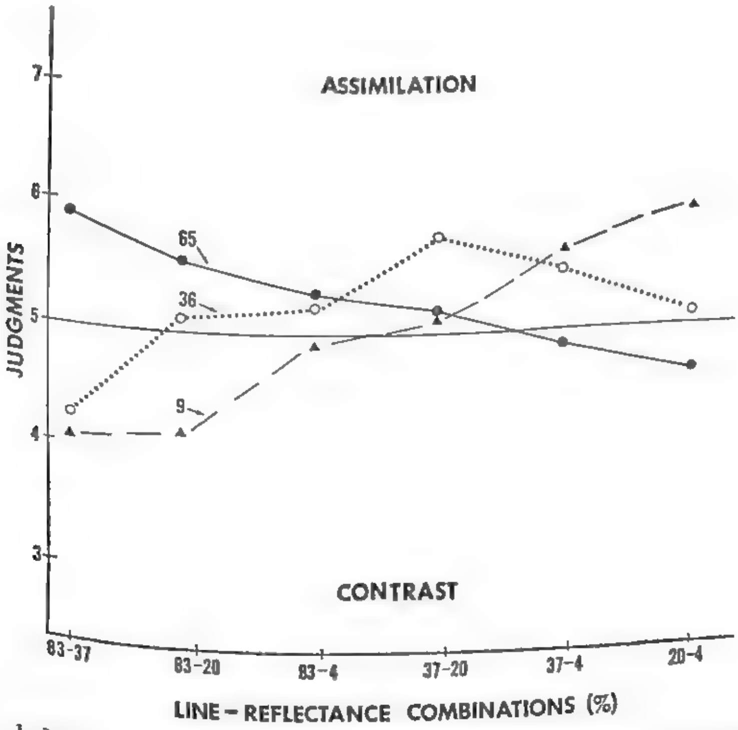


FIG. 1. MEAN JUDGMENTS CONFOUNDING ALL LINE AND SPACE WIDTHS: TECHNIQUE A

(The parameter for each curve is the reflectance of the intervening gray, *i.e.* 65%, 36%, and 9%.)

TABLE I

MEAN JUDGMENTS: TECHNIQUE A

Background (%)	Line × Space width (mm.)	Line-reflectance combination (%)					
		83–37	83–20	83–4	37–20	37–4	20–4
65	16 × 3	5.17	4.87	4.52	5.25	4.20	3.82
	16 × 9	5.25	5.67	4.57	5.37	4.67	4.10
	16 × 16	5.55	5.25	5.37	5.15	4.50	4.75
	9 × 3	4.95	4.50	4.35	4.84	3.92	4.45
	9 × 9	5.80	5.52	5.77	5.44	5.19	5.00
	9 × 16	5.62	5.35	5.42	5.40	5.27	4.87
	3 × 3	7.11	7.20	6.27	5.52	5.57	5.30
	3 × 9	6.57	6.87	6.90	5.57	5.65	5.07
	3 × 16	5.75	5.74	6.00	5.50	5.49	5.00
36	16 × 3	3.32	3.95	3.35	5.32	4.00	4.19
	16 × 9	3.94	4.34	4.45	5.57	4.91	4.42
	16 × 16	4.77	5.20	5.25	5.72	5.77	5.17
	9 × 3	3.79	4.25	5.02	5.35	5.70	4.72
	9 × 9	3.75	4.92	5.45	5.75	5.49	5.39
	9 × 16	4.29	5.30	5.77	5.72	5.62	5.55
	3 × 3	5.27	6.47	6.37	6.30	6.38	5.37
	3 × 9	4.84	6.10	6.00	6.17	6.40	5.57
	3 × 16	5.17	6.00	5.92	6.02	5.84	5.62
9	16 × 3	4.03	3.87	3.70	4.34	3.80	4.72
	16 × 9	4.15	3.59	3.75	4.95	4.65	5.17
	16 × 16	4.12	4.22	4.64	5.12	4.97	4.95
	9 × 3	3.94	3.82	3.87	4.70	5.27	5.92
	9 × 9	4.04	4.12	5.24	5.37	6.86	6.62
	9 × 16	4.49	4.47	6.22	5.47	6.07	6.14
	3 × 3	4.22	4.42	5.97	5.55	6.45	6.19
	3 × 9	4.27	4.27	5.77	5.80	6.45	6.19
	3 × 16	4.49	4.35	5.22	5.52	6.40	6.32

Note: $N = 10$. Values below 5 indicate contrast; those above 5 indicate assimilation.

These three response functions were a product of the reflectance differentials between the intervening gray and the overlaid line reflectances. With large differences between the reflectances of the lines and the ground, contrast resulted. When these reflectance differentials were smaller, assimilation appeared. Between these large and small reflectance differentials was an intermediate differential which produced responses of 'Equal.' In this case, differences in contiguous reflectances produced neither contrast nor assimilation. The three curves in Fig. 1 present the responses as a function of the line-reflectance combinations while confounding the line- and space-width combinations. Fig. 1 shows that the 65% ground reflectance yielded assimilation with line combinations of 83%–37%, 83%–20%, and 83%–4% reflectance. This is in accord with the differential-stimulation hypothesis in that each of these combinations contains the light (83%) line reflectance which is similar in reflectance to the ground (65%). Responses near 'Equal' were re-

ported when the intermediate line reflectance (37%–20% and 37%–4%) combinations were used. And following in expected order was the finding that the dark line combination (20%–4%) when used to overlay this light gray (65%) yielded contrast. The 9% intervening gray background produced responses complementary to those reported for the light (65%) background. Thus, the light line combinations (83%–37% and 83%–20%) on this dark (9%) ground yielded responses of contrast. The 83%–4% and 37%–20% line reflectances yielded responses of both assimilation and contrast, but neutral responses were most frequently reported. The line combinations of 37%–4% and 20%–4% reflectance produced responses of the greatest assimilation magnitude, but with wide lines, reports of contrast also occurred. The intermediate (36%) intervening gray produced responses between those of the light (65%) and the dark (9%) intervening grays. Responses of contrast were reported when this 36% gray was overlaid with the lightest line combination (83%–37%), responses of or near 'Equal' occurred when it was overlaid with the 83%–20% and 20%–4% line-reflectance combinations, and as the line-reflectance combinations of 83%–4%, 37%–20%, and 37%–4% approached the reflectance of the 36% ground, responses of assimilation were reported.

The data in Table I suggest certain generalizations. The first is that as line and space widths increased from 3 × 3 mm. to 16 × 16 mm., responses tended to change from assimilation to contrast. The exception is that when line and ground reflectances were similar in reflectance, or when small reflectance differentials existed, responses of assimilation were mentioned even with large line and space widths. The second generalization becomes apparent when one studies the relation of line reflectance to ground reflectance. The smaller the differential of line reflectance to background reflectance, the more responses of assimilation, or the less responses of contrast, occurred; the larger the reflectance differential between line and background, the more responses of contrast, or the less responses of assimilation, occurred; and with intermediate differentials in reflectance between line and ground, the more responses of 'Equal' or neutrality occurred.

Table II presents the statistical analysis of the findings with Technique A. It shows that the different line-reflectance combinations produced significant ($p < .001$) differences in response. However, it is also apparent that line reflectance must be considered in relation to background reflectance and line width, since the Back-

TABLE II

Analysis of Variance of Judgments: Technique A

Source	df	SS	MS	F
Individuals (O)	9	143.3387	15.9265	1.48
Backgrounds (B)	2	28.3680	14.1840	1.32
O × B	18	193.8460	10.7692	
Line-reflectance combinations (R)	5	86.2684	17.2537	16.13***
83%–37%, 83%–20%, and 83%–4% vs. 37%–20%, 37%–4%, and 20%–4%	1	51.3067	51.3067	47.98***
83%–37% vs. 83%–4%	1	28.6120	28.6120	26.76***
83%–37% and 83%–4% vs. 83%–20%	1	.0173	.0173	.02
37%–20% vs. 37%–4%	1	.2579	.2579	.24
37%–20% and 37%–4% vs. 20%–4%	1	6.0745	6.0745	5.68*
Line widths (L)	2	345.1386	172.5693	161.37***
Space widths (S)	2	49.5118	24.7559	23.15***
B × R	10	333.3067	33.3307	31.17***
B × L	4	21.2484	5.3121	4.97***
B × S	4	10.1311	2.5328	2.37*
R × L	10	59.7683	5.9768	5.59***
R × S	10	10.9000	1.0900	1.02
L × S	4	78.5306	19.6326	18.36***
B × R × L	20	35.9107	1.7955	1.68*
B × R × S	20	16.8298	.8415	.79
B × L × S	8	14.2128	1.7766	1.66
R × L × S	20	16.6481	.8324	.78
B × R × L × S	40	35.7838	.8946	.84
Error	1431	1530.3163	1.0694	
Total	1619	3010.0581		

* $p < .05$.
*** $p < .01$.

ground × Line Reflectance ($p < .001$), Line Reflectance × Line Width ($p < .001$), and Background × Line Reflectance × Line Width ($p < .05$) interactions all proved to be statistically significant. One must also add to these stimulus interactions, the significant contributions of line width ($p < .001$), space width ($p < .001$) and their significant interaction (Line Width × Space Width, $p < .001$), as well as the interactions of these stimulus dimensions with background reflectance (Background × Line Width, $p < .001$, and Background × Space Width, $p < .05$). The interpretation of these interactions in the statistical analysis should be viewed with caution, however, since the use of the nine-point rating scale may have erroneously produced nonadditivity in our statistical treatment. One cannot assume theoretically that the effects are not additive. We are now doing research on the use of a matching procedure in combination with the haploscopic viewing technique (see Diamond, footnote 2) to overcome the limitations of the rating scale.

It should be noted that with Technique A one must consider two

reflectance relationships operating simultaneously: the relation of the line reflectance to the ground reflectance and the relation of the two different line reflectances within any given line-reflectance combination. It is evident from the results shown in Fig. 1 that a combination of two lines each similar in reflectance to the ground reflectance produced a greater degree of assimilation than a combination of one line of similar reflectance and one line of dissimilar reflectance to the ground. The converse also held for responses of contrast in that a combination of two lines, each of dissimilar reflectance to the ground, produced more contrast than a combination of one line of dissimilar reflectance and one of similar reflectance to the ground. For example, the 83%–37% line-reflectance combination overlaid on the 65% ground produced responses of assimilation of a greater magnitude than did the 83%–4% line-reflectance combination on the same ground.

This finding raises a question of methodological importance, *i.e.* whether only one of these line reflectances in the combination actually accounted for the results. Or, more specifically, whether there was an artifact which might have confounded the actual response when the responses to each of the lined sides of the stimulus card were averaged. The answer depends upon the conceptual framework from which one considers the question. If one views the situation as containing two separate stimuli (the two sides of the stimulus card) and hence two different responses (one response to each of the lined sides), then the responses would have to be considered separately. One may have assimilation reported for one side of the stimulus card and contrast reported for the other side, as did occur for (only) 9 of the 162 stimulus cards. However, since assimilation and contrast have been found to be but different zones of a total bipolar response function, one must consider the total viewing situation as the stimulus to which *S* was responding. If one views the situation in this manner, it would be the average of the responses to the two sides of the stimulus card that should be assigned to the given combination of stimulus factors. Thus one side is not a separate stimulus complex producing responses of one kind and the other side another separate complex producing responses of a different kind, but the total is a stimulus complex producing one response. Thus too, Technique A did not therefore alter in any significant manner the results when they are considered as but one bipolar response function. Evidence to support this contention is found by the close relationship of the findings with Technique A to those with Technique B,

which was designed to overcome the methodological difficulty encountered with the use of two sides of different line reflectances for each stimulus card.

Technique B. The results with Technique B (in which the ruled background was compared with a homogeneous background of the same reflectance) substantiate the findings with Technique A and add more evidence for the validity of the stimulation-differential and area-luminance hypotheses. Table III shows the mean responses as a function of all stimuli with Technique B. From these results as well as from those plotted in Fig. 2 (which confounds line and space widths), it is apparent that the same bipolar response function was found with Technique B as with A. Responses of assimilation and contrast with both techniques were a function of the reflectance differentials (neglecting, for the moment, line and space factors). Fig. 2 shows that large differences between the line and ground produced responses of assimilation. The intermediate reflectance dif-

TABLE III

MEAN JUDGMENTS: TECHNIQUE B

Background (%)	Line × Space width (mm.)	Line reflectance (%)			
		83	37	20	4
65	16 × 3	5.00	4.60	4.75	3.85
	16 × 9	4.80	4.85	4.35	3.55
	16 × 16	4.95	4.85	4.30	4.00
	9 × 3	5.25	4.80	4.75	4.35
	9 × 9	5.50	4.95	4.80	4.10
	9 × 16	5.10	4.70	4.55	3.90
	3 × 3	5.45	5.35	5.25	5.30
	3 × 9	5.35	5.20	4.90	4.30
	3 × 16	4.80	4.80	4.55	4.40
36	16 × 3	4.60	5.60	5.05	4.65
	16 × 9	4.10	4.65	4.60	4.35
	16 × 16	4.35	5.25	4.75	4.40
	9 × 3	4.45	5.90	5.50	4.60
	9 × 9	4.40	4.85	5.30	4.75
	9 × 16	4.45	4.95	5.65	5.05
	3 × 3	5.50	4.95	5.75	5.30
	3 × 9	4.95	4.95	5.70	5.40
	3 × 16	4.75	5.00	5.60	5.25
9	16 × 3	3.65	4.85	4.95	5.10
	16 × 9	3.60	4.15	4.70	5.25
	16 × 16	3.10	4.15	4.10	5.85
	9 × 3	4.10	5.25	5.50	5.30
	9 × 9	3.45	4.45	4.60	5.85
	9 × 16	3.45	4.20	4.35	5.35
	3 × 3	5.05	5.75	6.00	5.60
	3 × 9	3.85	4.50	5.00	6.05
	3 × 16	3.90	4.65	4.45	6.10

Note: $N = 10$. Values below 5 indicate contrast; those above 5 indicate assimilation.

ferentials yielded neither contrast nor assimilation. Comparing the responses made to each background reflectance, one can observe in Fig. 2 three bipolar response functions, one for each of the three backgrounds, as with Technique A and shown in Fig. 1. Fig. 2 shows that the 65% background reflectance yielded a small amount of response assimilation when overlaid with the lightest line reflectance (83%). As the lines decreased in reflectance (37%, 20%, and 4%), and hence the difference in reflectance between the 65% background and the line reflectances increased, reports of contrast occurred. A complementary response function was produced with the 9% background. Responses of assimilation were reported when this dark (9%) background was overlaid with the darkest (4%) line reflectance, and responses of contrast occurred in increasing magnitude as the overlaid lines increased in reflectance from 20% to 37% to 83%. Thus, as the line reflectance became lighter in relation to this dark ground, larger differences in stimulation occurred and responses of contrast resulted. Fig. 2 also shows that with the 36% background, both the highest (83%) and the lowest (4%) line reflectances pro-

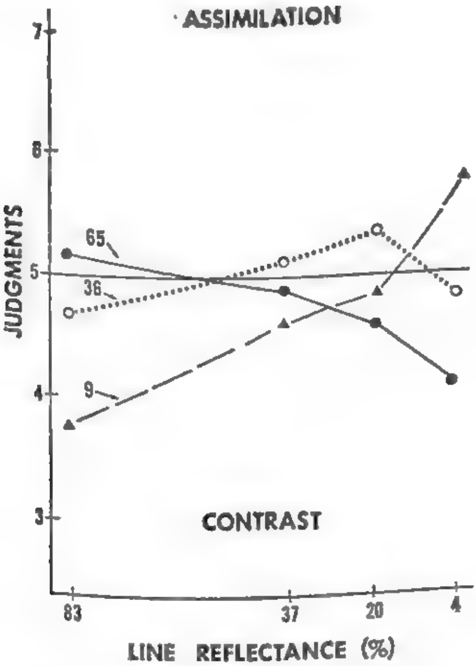


FIG. 2. MEAN JUDGMENTS CONFOUNDING ALL LINE AND SPACE WIDTHS: TECHNIQUE B

(The parameter for each curve is the reflectance of the intervening gray, *i.e.* 65%, 36%, and 9%.)

stimulation yielded responses of contrast, small differentials in stimulation yielded responses of assimilation, and intermediate differentials in stimulation yielded responses of neither contrast nor assimilation. Assimilation, contrast, and no effect of the lines on the backgrounds can thus be seen as the orderly consequences of different stimulus conditions which produce varied stimulation differentials.

It is apparent that the present findings also support the area-luminance hypothesis. However, increases in area of a given reflectance do not act like simple increases in luminance of fixed areas, as Helson has previously pointed out.[14] Thus, with intervening gray of 65% reflectance and 3-mm. lines of 83% reflectance, the result is assimilation. If the 3-mm. lines and spaces are increased in area by a factor of 5.3, to 16 mm., the result is responses of contrast. On the other hand, a gray of 9% reflectance with 3- or 16-mm. lines of 4% reflectance elicits only responses of assimilation. The reflectance ratio of the 65% gray to the 83% line is only 1.27:1, whereas the 9% gray to the 4% line has a ratio of 2.25:1. From the ratios of reflectances one would predict that the larger ratio should yield responses of contrast. However, from the present findings it is apparent that smaller changes in area of the higher reflectances yield larger changes than do equivalent changes in the lower reflectances. Blackwell's report that liminal contrast is more readily achieved at higher levels of field intensity,[15] Beck's finding that contrast is reported when the reflectance of the overlaid lines or circles is above that of the ground and that assimilation is reported when the reflectance is below that of the ground,[16] and the present study's indication that higher reflectances appear to increase in stimulation intensity at a greater rate than lower reflectances, all thus suggest this hypothesis: At high overall levels of stimulation (both line and ground of high reflectance), smaller differentials of stimulation will produce more responses of contrast than occur at low overall levels of stimulation (both line and ground of low reflectance). This hypothesis can be tested. If it is valid, with equal stimulation differentials (assuming the area-luminance function has been determined for each reflectance level), the condition resulting in the lower overall intensity (adaptation level) should yield responses of

[14] Helson, *loc. cit.*

[15] H. R. Blackwell, Contrast thresholds of the human eye, *J. opt. Soc. Amer.*, 36, 1946, 624–643.

[16] J. Beck, Contrast and assimilation in brightness judgments, *Psychon. Sci.*, 1, 1966, 342-344.

less contrast (more assimilation) than the condition resulting in the higher overall intensity (adaptation level). Further, if it is valid, with equal stimulation differentials, responses should be a function of the adaptation levels, the higher adaptation levels yielding more responses of contrast (less assimilation) than the intermediate levels and the lower adaptation levels yielding more responses of assimilation (less contrast) than the high or intermediate levels.

Several methodological notes conclude the discussion. The technique previously employed by Helson and his co-workers[17] and similarly employed in this study is lacking in several respects. As Hurvich has pointed out, the use of the split-card comparisons confounds what *S* views on each side of the card, and since his reaction is summed, the net result can be neither contrast nor assimilation when in fact each side may be producing the opposite effect.[18] Hurvich also suggested that since the lines of different reflectance change effects at different rates this asymmetry of effect is also confounded. A remedy for these methodological problems as well as for the problems usually associated with the use of a nine-point scale is the use of a matching procedure, as Hurvich has suggested.[19] Steger has further suggested that the haploscopic viewing situation may serve as a useful technique in the matching procedure.[20] Thus, the method of adjustment coupled with a standard to one eye and a variable to the other may, hopefully, overcome the methodological shortcomings of the present procedure. This new approach is currently being developed in our laboratory.

Summary

Two techniques were used to investigate the stimulus interactions in visual lightness assimilation and contrast in relation to Helson's differential-stimulation hypothesis. Twenty *S*s each made 54 comparisons four times. The stimuli were 7 × 11 in. cards of 65%, 36%, and 9% reflectance on which were overlaid vertical lines of 83%, 37%, 20%, and 4% reflectance. Combinations of lines and spaces of 3, 9, and 16 mm. were used. With Technique A, vertical lines of different reflectances but of equivalent line and space widths

[17] Helson and Rohles, *loc. cit.*; Helson and Joy, *loc. cit.*
[18] L. M. Hurvich, Contributions to color-discrimination theory: Review, summary, and discussion, *J. opt. Soc. Amer.*, 53, 1963, 196–201.
[19] Hurvich, *loc. cit.*
[20] Steger, *loc. cit.*

were overlaid on each half of each card. With Technique B, only one side of each card was overlaid with lines; the other side was plain. The results with both techniques support the differential-stimulation hypothesis. One new finding emerged: to produce contrast, larger differentials in reflectance were more necessary at the low end of the reflectance continuum than at the high end.

STIMULUS FACTORS IN OBSERVER JUDGMENT OF SOCIAL INTERACTION: FACIAL EXPRESSION AND MOTION PATTERN

By Stephen Thayer and William Schiff,
The City College of the City University of New York

Judgment of interpersonal behavior plays a critical role in everyday life. But although psychologists have long recognized the importance of social cues governing such judgments, there has been little systematic research on observer judgments of human social interaction and relationships.[1] Two variables that apparently do influence an observer's judgment about the nature of a social relationship are the emotional states (as indicated in part by facial expressions) of the participants and the movements they make relative to one another.[2] In an early effort to experimentally explore stimulus factors in the judgment of social relationships, Heider and Simmel used motion pictures of faceless, interacting geometric figures.[3] Their study examined the influence of spatiotemporal stimuli on social perception by presenting figures whose movements were sufficient conditions for their perception as 'people.' *Ss* readily made fairly consistent judgments about the apparent emotions, motives, and social relationships of the figures when asked to describe the events in the films. Building on this work, Michotte emphasized the all-encompassing role of movement in affecting judgments of social relationships.[4] According to Michotte, every factor that influenced a perceived social interaction was subordinate to the overwhelming factor of movement. He wrote: "In ordinary life, the specifying factors—gestures, facial expressions, speech—are innumerable and can be differentiated by an infinity of nuances. But they

[1] A. Bandura and R. H. Walters, *Social Learning and Personality Development*, 1963, 114; R. Q. Bell, Developmental psychology, in P. R. Farnsworth, Olga McNemar, and Q. McNemar (eds.), *Annu. Rev. Psychol.*, 16, 1965, 26; J. J. Gibson, Theories of perception, in W. Dennis (ed.), *Current Trends in Psychological Theory*, 1951, 95–96.

[2] M. G. Cline, The influence of social context on the perception of faces, *J. Pers.*, 25, 1956, 142–158; A. Michotte, The emotions as functional connections, in M. L. Reymert (ed.), *Feelings and Emotions*, 1950, 114–126.

[3] F. Heider and Marianne Simmel, A study of apparent behavior, this Journal, 57, 1944, 243–259.

[4] Michotte, *loc. cit.*

are all *additional refinements* compared with the key factors, which are the simple kinetic structures."[5] More explicitly, he argued that motor reactions corresponding to the "integrative" emotions (*e.g.* sympathy, friendship, love) usually result in ". . . union, possession . . . , approach . . . positioning oneself close to another . . . contact more or less pronounced, more or less prolonged. . . ."[6] Similarly, he maintained that the characteristic motor reactions of the "segregative" emotions (*e.g.* antipathy, disgust, hate, fear) "naturally present the opposite attributes. We move away from the person we do not like."[7] Since Michotte's work, several researchers have followed the technique of using faceless interacting animated figures, and each has found fairly consistent judgments about the nature of the social relationships involved, for different populations, while using different motion patterns.[8]

Everyday experience suggests that facial expressions of interacting parties will also influence an observer's judgment about their relationship. Cline's study of the influence of social context on the perception of faces shows that the respective facial expressions of a pair of people play an important part in judgments of their relationship.[9] Cline found that the feelings and motives attributed to one figure were often influenced by the facial expression of the figure with which it was paired. The present study was an attempt to extend the systematic exploration of stimulus factors affecting perception and judgment of social interaction, for it appears that a great deal of information about the nature of the social relationships involved may have been lost with the use of interacting *faceless* figures. Facial expression may be more or less important for judgments of certain motion patterns as opposed to others. That is, facial expression may complement or contradict inferences drawn from particular social encounters. What effect this may have on observer judgment is presently unknown. The present study was concerned with the hypothesis that the stimulus information pro-

[5] Michotte, *op. cit.*, 122.
[6] Michotte, *op. cit.*, 117.
[7] Michotte, *op. cit.*, 117.
[8] L. Buck and S. L. Kates, Perceptual categorizations of love and anger cues in schizophrenics, *J. abnorm. soc. Psychol.*, 67, 1963, 480–490; Dorothy C. Krugman, Structural and functional determinants in the perception of human-like behavior, unpublished doctoral dissertation, Columbia University, 1950; J. Marek, Information, perception, and social context: I. Simple level of perceptual response, *Hum. Relat.*, 16, 1963, 209–231; R. E. Shor, Effect of preinformation upon human characteristics attributed to animated geometric figures, *J. abnorm. soc. Psychol.*, 54, 1957, 124–126.
[9] Cline, *loc. cit.*

vided by motion pattern alone would influence observer judgment of simple social relationships but that apparent emotional state, as indicated by facial expression, should modify such judgments when the apparent emotional state was in contradiction with motion pattern, even when motion pattern was made highly salient.

Method

Stimulus materials. In a preliminary study, a set of 10 faces, plus a control 'nonsense' face, was drawn on 1⅝-in. diameter discs, photographed on 35-mm. Kodachrome film,[10] and shown in random order to 18 seniors at CCNY. The set of faces was similar to that tested extensively by Harrison[11] and Strattner,[12] although independently conceived. The *Ss* were instructed to check one of 14 one-word descriptions of facial expressions (affective states) most appropriate to each expression and to indicate the intensity of each expression on a five-point scale ranging from 'slightly' to 'extremely.' The faces, one-word descriptions, and *Ss* responses are shown in Table I, which reveals that most *Ss* agreed in their responses to several expressions. Since the data showed that the Angry$_2$ and Happy faces were among those most clearly differentiated from each other and from the Neutral expression, they were chosen for use, along with the Control and Neutral faces. Happy and Angry$_2$ also represent two extremes of an expressive continuum well suited for use in sequences depicting social interaction, with Neutral as the continuum midpoint.

A set of 16 animated films was then produced by photographing pairs of these Control, Neutral, Happy, and Angry faces on a lightly ruled white metal background. The films were 16 mm. and were made with a single-frame exposure technique. The faces were moved ¾ in. every two frames. Each film was 200 frames long, providing a projection time of 12.5 sec. at 16 frames per sec. Three basic motion patterns plus a control sequence with no motion were photographed. Each pair of faces was filmed in each of the patterns. In addition to the linear movement of the faces across the field, they were alternatively turned 12° to the left and right each time they were moved, giving an animated appearance similar to the movement in the films used by Heider and Simmel[13] and Krugman.[14] In the first motion pattern (Passive Receiver, PaR), Face B remained motionless during A's approach. In the second pattern (Approaching Receiver, ApR), B approached A after A began approaching B. In the third pattern (Avoiding Receiver, AvR), B moved away from A after A began approaching B. All motion was of the same velocity. These patterns provided "integrative" and "segregative" relationships according to

[10] The authors are indebted to Kenneth Brown for the production of these slides.

[11] R. P. Harrison, Pictic analysis: Toward a vocabulary and syntax for the pictorial code; with research on facial communication, unpublished doctoral dissertation, Michigan State University, 1964.

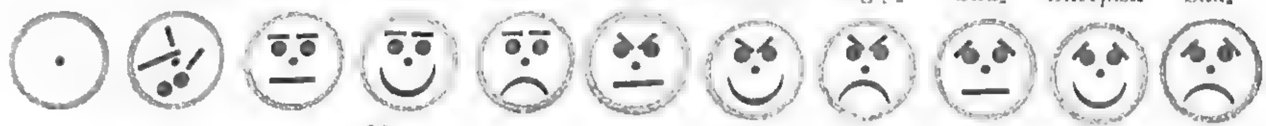
[12] Mary J. Strattner, A developmental study of young children's perception of affect in drawing of faces and postures, unpublished doctoral dissertation, Cornell University, 1963.

[13] Heider and Simmel, *loc. cit.*

[14] Krugman, *loc. cit.*

TABLE I

Percent (Rounded) Responses to Faces in Each Descriptive Response Category and Modal Intensity Rating of Predominant Judgment (In Parentheses)



Descriptions		Blank	Control	Neutral	Happy	Sad_1	$Angry_1$	Fiendish	$Angry_2$	Sad_2	Happy-Sheepish	Sad_3
Elated	1				03			03			03	
Happy	2		11	33 (2)	81 (4)						36 (3)	
Neutral	3	89 (5)	38 (4)	39 (3)						06		
Sad	4			08		42 (3)			03	50 (3)		61 (5)
Angry	5		22 (2)			13	58 (4)		53 (5)			
Furious	6						17	03	33 (4)			
Amused	7			03	08		03				19 (3)	
Sheepish	8		06	03		03				03	25 (3)	
Mischievous	9		06				11	16			06	
Fiendish	10				08		03	75 (5)	11	03		
Depressed	11					17	03			08		24
Apprehensive	12		06	06		19	03			22 (3)	12	06
Afraid	13			03		03				03		06
Horrified	14	03				03						
Other	15	06	11	03				03		06		03

Michotte's classification.[15] In all three patterns a 'contact' sequence was included which lasted 32 frames and occurred at the same spatial and temporal point in each film. During these sequences, Face A was moved 1½ in. toward B, providing the appearance of A touching or striking B eight times in succession while B remained motionless. Following the contact sequence, B left the field in the direction opposite to that from which A approached. In the Control sequence (No Motion, NoM), A and B remained motionless for the entire film, at a distance midway between the starting positions of A and B in the motion sequences. This provided a partial control for proximity. In all sequences, the faces were first identified by the labels A and B. After 68 frames, the labels disappeared. The complete script is shown in Table II. The four motion patterns and four facial expressions were combined to yield a 4 × 4 design.

Subjects. The *Ss* were 345 volunteers from introductory psychology classes at CCNY. Five *Ss'* data were discarded from the analyses due to incomplete or unreadable response sheets, leaving a total of 340 *Ss*. Of these, 203 were men, 137 were women. Ages ranged from 16 to 50 years and averaged 19.5 years.

TABLE II

'Script' for Three Motion Patterns and One Control Sequence

	Sequence	No. frames
Passive Receiver (PaR)	A and B identified with letters	68
	Letters disappear; no motion	16
	A moves toward B	28
	A bumps B eight times	32
	Pause; no motion	2
	B moves away from A and off screen	22
	A alone on screen	32
	Total	200
Approaching Receiver (ApR)	A and B identified with letters	68
	Letters disappear; no motion	16
	A moves toward B only	20
	B moves toward A at same rate	8
	A bumps B eight times	32
	Pause; no motion	2
	B moves away from A and off screen	22
	A alone on screen	32
	Total	200
Avoiding Receiver (AvR)	A and B identified with letters	68
	Letters disappear; no motion	16
	A moves toward B only	20
	B moves away from A at same rate	8
	A bumps B eight times	32
	Pause; no motion	2
	B moves away from A and off screen	22
	A alone on screen	32
	Total	200
Control, No Motion (NoM)	A and B identified with letters	68
	Letters disappear; no motion	132
	Total	200

[15] Michotte, *op. cit.*, 117.

Dependent measures. A number of questions were included on response sheets to assess the effects of the various components of the stimulus materials. Question 1 was designed to test *S*'s attention to the film: "Which of these were actually in the movie? Put an A under face A and a B under face B [followed by drawings of the faces used, plus the Sad₃ face]." Question 2 was open-ended, to obtain qualitative descriptions of the film, free of possible suggestions provided by the labeled scales which followed: "Briefly describe what happened." Question 3 was the primary dependent measure and was based on Michotte's distinction between "integrative" (friendly) and "segregative" (hostile) social relationships: "Check the place on the following scale that best tells what went on overall [followed by a seven-point friendliness/hostility scale, labeled: Extremely friendly or hostile; Moderately friendly or hostile; Slightly friendly or hostile; Neutral]." Questions 4 and 5 were included to verify the *E*s' and pilot *S*s' labels for the facial expressions: "How did A look? How strong was A's look?" and "How did B look? How strong was B's look? [followed by *Looks* Happy, Sad, Angry, Neutral, Afraid, Love, Could Not Tell, Other; *Strengths* Extremely, Moderately, Somewhat, Slightly]."

Procedure. Each group of 8–15 *S*s was seated 5–10 ft. from a screen and was shown only one film. Total projected field size was 30 × 40 in., and projected face diameters were 2¼ in. *S*s were randomly assigned to conditions and instructed:

> I'm going to show you a very brief film. Afterwards, you'll be asked to write answers to some questions about what you saw. As soon as the film is over, I'll distribute these folded response sheets. Please answer the first two questions without looking at the next three (folded under). Then turn your papers over and answer the next three. When you've finished, please write your age and sex in the spaces provided. Don't change your other answers. Now please look carefully at the screen. You'll see figure A on the right and figure B on the left. Raise your hand if there are any questions.

*S*s were told the purpose of the study after completing the response sheets and were asked not to reveal the nature of the study to classmates.[10]

Results

Table III presents a summary of a 4 × 4 analysis of variance performed on friendliness/hostility (f/h) judgments of encounters. This analysis shows that both facial expression and motion pattern significantly influenced f/h judgments. To further examine the main effects, *t* tests were performed on the variations within the two major variables. Table IV presents the means and *SD*s. Table V presents a summary of *t*-test results for facial expression and motion pattern; it reveals that all facial expressions yielded significantly different f/h judgments, with the exception of Control and Neutral faces. The three motion patterns did not differ significantly from each other, although each was significantly different from the

[10] Five *S*s who knew about the purpose of the study were eliminated before film screenings.

TABLE III

Analysis of Variance of f/h Judgments

Source	*SS*	*df*	*MS*	*F*
Facial expression	291.73	3	97.24	37.54*
Motion pattern	108.07	3	36.02	13.91*
Facial expression × motion pattern	32.40	9	3.60	1.39
Within	837.85	324	2.59	

* $p < .001$.

NoM condition. Table VI presents the f/h response distribution, in percentages, for each encounter, classified by motion pattern and facial expression.

With respect to the qualitative data, responses to the request "Briefly describe what happened" show the different ways in which *Ss* integrated facial expression and motion cues. The two most divergent f/h response distributions were obtained with the Happy as opposed to the Angry faces (Table VI). Regardless of the motion pattern used, f/h judgments of encounters containing Angry faces tended to fall at the hostile end of the scale. Considering Happy faces, however, it can be seen that f/h responses tended to be more equally divided between friendly and hostile judgments. That is to say, motion pattern did not combine similarly with Angry and Happy faces but had a greater impact with Happy faces. Following are typical hostile and friendly descriptions *Ss* gave of encounters containing Happy faces in the different motion patterns.

(*1*) *Hostile judgments*. In the main, for the AvR pattern, hostile intrepretations characterized the contact between figures as A having forcefully driven off B. A's actions (*e.g.* "kicked," "strikes," "knocked," "attacked") were reported as the immediate cause for B's leaving. Hostile interpretations for the ApR pattern, however,

TABLE IV

Means and *SD*s of f/h Judgments

	Facial expression			
	Control	Neutral	Happy	Angry
Mean	5.13	4.86	3.45	6.04
SD	1.65	1.77	2.04	1.23
	Motion pattern			
	AvR	ApR	PaR	NoM
Mean	5.36	5.07	5.39	3.85
SD	1.75	2.14	1.80	1.50

TABLE V

SUMMARY OF t-TEST RESULTS

Comparison	t
Facial expression	
Control-Neutral	1.03
Control-Happy	5.88*
Control-Angry	4.02*
Neutral-Happy	4.82*
Neutral-Angry	5.02*
Happy-Angry	9.99*
Motion pattern	
NoM-AvR	5.97*
NoM-PaR	5.94*
NoM-ApR	4.28*

Note: All remaining ts < 1.0.
* $p < .001$.

were not as violent as the AvR pattern, although contact was still always mentioned. Typical responses included: "B nudges A, A didn't respond and B left," "B pushed," and "B bumped A." This reversal of A's and B's roles apparently occurred because in the ApR pattern B moved toward A. Hostile interpretations for the PaR pattern were almost equally divided between A directly and

TABLE VI

DISTRIBUTION OF f/h RESPONSES

Facial expression	Motion pattern	Scale categories 1	2	3	4	5	6	7
Control	ApR	8.3	0.0	4.2	4.2	12.5	33.3	37.5
	PaR	5.0	5.0	0.0	5.0	25.0	35.0	25.0
	AvR	0.0	4.5	4.5	0.0	36.4	40.9	13.6
	NoM	0.0	10.5	10.5	68.4	5.3	5.3	0.0
Neutral	ApR	8.3	0.0	4.2	8.3	20.8	25.0	33.3
	PaR	5.0	5.0	5.0	10.0	0.0	40.0	35.0
	AvR	4.8	4.8	9.5	4.8	28.6	38.1	9.5
	NoM	10.0	5.0	25.0	60.0	0.0	0.0	0.0
Happy	ApR	36.4	22.7	4.5	9.1	9.1	9.1	9.1
	PaR	23.8	4.8	4.8	19.0	14.3	19.0	14.3
	AvR	22.7	13.6	9.1	0.0	13.6	27.3	13.6
	NoM	9.1	36.4	31.8	18.1	4.5	0.0	0.0
Angry	ApR	0.0	4.5	4.5	0.0	4.5	27.3	59.1
	PaR	0.0	4.5	4.5	0.0	0.0	40.9	50.0
	AvR	0.0	0.0	0.0	4.8	14.3	42.8	38.1
	NoM	0.0	0.0	5.0	20.0	15.0	30.0	30.0

Note: 1 = extremely friendly, 7 = extremely hostile. Distribution shown in percentages.

forcefully driving B off and A attacking B, but B leaving of his own volition. Typical responses were: "A knocked B off the screen," "A hit B a few times and then B moved out," and "A attacked B a series of times, then B withdrew." Thus, the combination of the Happy face and the integrative motion pattern (ApR) was seen as the least aggressive encounter, and B was not seen as a 'victim.' The combination of the Happy face and the segregative motion pattern (AvR) was seen as a more aggressive encounter, and B was seen as the 'victim.' Lastly, encounters containing the combination of the Happy face and the motion pattern midway between segregative and integrative (PaR) led to judgments that the encounter was aggressive, but B was not consistently seen as the 'victim.' Considered as a whole, the qualitative data confirmed the results obtained with f/h scaling. Specifically, although integrative and segregative motion patterns played a role in determining the appearance of an encounter, it was hardly the primary role proposed by Michotte.

(2) Friendly judgments. Friendly interpretations for each of the three motion patterns with the Happy faces also had a different nuance. In the AvR pattern, the encounter was described as involving gentle contact or no contact. Typical responses were: "A says hello to B and B left," "A touched B and B left," and "A joined B" (no contact mentioned). Responses for the ApR pattern were characterized by either affectionate contact or nonaffectionate contact without violence. Typical descriptions included: "A tried to make contact with B," "B kissed A," "A cuddled B," or "A began bouncing against B." Friendly interpretations of the PaR pattern involved playful actions with an aloof recipient: "A played with B," "A pecked at B a few times, then B ambled away," "A made friendly advances; B while not rejecting him, danced playfully out of his reach." The more affectionate qualitative descriptions reported for the ApR and, to a lesser extent, for the PaR as compared with the AvR pattern, again provided confirmation of the role, albeit minor, played by integrative and segregative patterns in the modification of observer judgment.

Identification accuracy. Lastly, *S*s were rather accurate in identifying the faces used in the films. Almost 80% of the *S*s correctly identified both faces, A and B. However, over 52% of the *S*s viewing Angry encounters made errors. Whether this result is due to stimulus-procedural or *S* factors is presently unknown.

Discussion

The results clearly show that the stimulus information provided by motion pattern per se is insufficient to account for observer judgment of certain social encounters. Information provided by facial expression of the interacting characters may significantly influence f/h judgments of even such relatively simple encounters as those used in the present study. Michotte's point that "facial expressions . . . are *additional refinements* compared with the key factors, which are the simple kinetic structures,"[17] appears in need of qualification. Michotte's notion that "integrative" and "segregative" movements are the *primary* cues for observer judgment[18] is challenged by the present results, for the same motion pattern often led to significantly different judgments with some faces. For example, the Happy-face encounters were judged significantly more friendly than the Angry-face encounters, regardless of motion pattern. Thus, rather than being "additional refinements," facial expression may function as critical stimulus information for the judgment of an encounter.

The presence and degree of conflicting stimulus information from motion pattern and facial expression may account for the differential results obtained with the Happy and Angry faces, especially the great variability in f/h judgments obtained with the Happy faces. A striking finding was that even with Happy faces, a large number of *S*s (44% for Scale Categories 5, 6, 7) judged the encounter as hostile when combined with motion (see Table VI). Since this finding was *not* corroborated with Happy faces in the NoM condition (one hostile response in Category 5), one must conclude that some part or parts of the motion pattern carried information of hostility. If this is true, then in the case of Angry faces, information of hostility may have been redundantly carried by both motion pattern and facial expression. In the case of Happy faces, however, conflicting information was carried by the two major stimulus sources: facial expression and motion. Considering the qualitative data as well as the f/h scaling data, some *S*s apparently resolved this conflict in favor of the facial expression, reinterpreted the motion pattern accordingly, and judged the encounter as friendly. Others apparently resolved the conflicting cues in favor of the motion pattern by either discounting the Happy facial expres-

17 Michotte, *op. cit.*, 122.
18 Michotte, *op. cit.*, 117.

sion or weighting it less, and judged the encounter as hostile. Since there was no cue conflict in sequences containing Happy faces with no motion, the response scatter found with other motion patterns was not present. *S*s used the only information they had (facial expression) to make a judgment.

The effect of introducing motion as a possibly conflicting cue is reflected by encounter judgments for the different motion patterns. Table VI shows that one cue, Happy faces, whose effects are made clear in the NoM pattern, were judged overwhelmingly friendly (Categories 1, 2, 3; over 77%). The potentially conflicting cue, the different motion patterns, whose effects are made clear with the Control faces, shows that these patterns were judged overwhelmingly hostile (PaR, ApR, AvR; Categories 5, 6, 7; over 86%). Therefore, the stimulus components of motion pattern and facial expression (Happy) were weighted to yield extremely different f/h judgments. An averaging or consistency model[19] or a summation model[20] would predict that most *S*s would respond in the nonextreme Categories 3, 4, 5, when these two conflicting cues were combined in one compound event. On the other hand, judgments of encounters containing Angry faces, in contrast to Happy faces, show little f/h variability between motion patterns. Apparently the combination of an Angry face with *any* motion pattern did not provide a cue-conflict situation but, rather, a redundantly hostile cue situation. Judgments of the Angry face combined with any one of the three motion patterns yielded a greater proportion of extremely hostile ratings (Categories 6, 7) than were obtained with either one of the two constituent elements alone. Thus, these findings for redundant stimulus cues may be viewed as support for what has been called an "enhancement of contrast" phenomenon where compound stimuli of high mean scale values receive judgments which are higher than the mean scale value of the component elements.[21]

One possible explanation for the discrepancy between the results of the present study (specifically Happy faces, AvR pattern) and

[19] C. E. Osgood and P. H. Tannenbaum, The principle of congruity in the prediction of attitude change, *Psychol. Rev.*, 62, 1955, 42–55.

[20] M. Fishbein and Rhoda Hunter, Summation versus balance in attitude organization and change, *J. abnorm. soc. Psychol.*, 69, 1964, 505–510.

[21] Jean S. Kerrick, The effect of relevant and nonrelevant sources on attitude change, *J. soc. Psychol.*, 47, 1958, 15–20; M. Manis, T. C. Gleason, and R. M. Dawes, Evaluation of complex social stimuli, *J. pers. soc. Psychol.*, 3, 1966, 404–419; W. Weiss, Scale judgments of triplets of opinion statements, *J. abnorm. soc. Psychol.*, 66, 1963, 471–479; R. H. Willis, Stimulus pooling and social perceptions, *J. abnorm. soc. Psychol.*, 60, 1960, 365–373.

studies involving explicit tests of stimulus-pooling models, may lie in the materials or procedures employed. In earlier studies, when conflicting information was given, it was presented sequentially, in writing or verbally.[22] This may have made it possible for judges to continually revise their judgments as congruent and incongruent information appeared; *i.e.* they were not confronted with simultaneously contradictory information but may have had a developing frame of reference within which to weigh and classify incoming information. Such a procedure would seem to enhance the likelihood of a summation or averaging effect. In the present study, however, conflicting information reached the *S* at the same time. Given conflicting cues presented simultaneously, *Ss'* extreme ratings may be accounted for by describing the *Ss* as: Discounters, who ignored either face or motion cues; Distorters, who perceived the weaker cue in line with the stronger; or Reinterpreters, who reinterpreted the weaker cue in line with the stronger. These ways of handling conflicting stimulus information point up differences in psychological processes evoked by the procedures of the pooling-model studies and those of the present study. Sequential presentation of verbal or written information allows for classification and 'storage,' since the messages are discrete wholes, sufficiently spaced in time to permit classification and hypothesis testing. Such procedures are appropriate for the study of cognition or thinking. In the present study, the simultaneous presentation of contradictory stimulus information did not necessarily allow for processing of the elements since the encounter was most likely experienced as a unit. Although such a 'global stimulus' could be analyzed into components, it is unlikely that *Ss* did so at the moment of observation. Thus, while *Ss* apparently responded differently to the discrepant information, they did not necessarily do so after the fashion of *Ss* testing hypotheses during sequentially presented information. More likely, our *Ss* were engaged in differential sampling of stimulus information, as opposed to differential decoding, encoding, retrieval, weighting, integration, and hypothesis testing. (The Reinterpreters may be an exception.)

[22] S. E. Asch, Forming impressions of personality, *J. abnorm. soc. Psychol.*, 41, 1946, 258–290; J. S. Bruner, D. Shapiro, and R. Tagiuri, The meaning of traits in isolation and combination, in R. Tagiuri and L. Petrullo (eds.), *Person Perception and Interpersonal Behavior*, 1958, 277–288; A. S. Luchins, Primacy-recency in impression formation, in C. I. Hovland, *et al.*, *The Order of Presentation in Persuasion*, 1957, 33–62; Weiss, *loc. cit.*; Fishbein and Hunter, *loc. cit.*; Marjorie H. Richey, Lucille McClelland, and A. M. Shimkunas, Relative influence of positive and negative information in impression formation and persistence, *J. pers. soc. Psychol.*, 6, 1967, 322–327.

In conclusion, the present study shows that the use of animated films with precisely controlled stimuli is a promising technique for the exploration of observer judgment of social behavior. Films allow the introduction and systematic manipulation of almost every variable (in particular, kinetic cues) that may contribute to judgment of social interaction.

Summary

Facial expression and relative movements of pairs of interacting figures were combined in a series of brief motion pictures to study their effects on observer judgments of the friendliness/hostility of the encounter. Facial expression and motion pattern were both significant factors affecting judgments of friendliness and hostility. Judgments depended upon the congruence of facial expression and motion pattern, for congruent combinations resulted in judgments more extreme than for either component alone, while incongruent combinations of conflicting cues failed to produce averaging effects. Averaging and summation models of stimulus-pooling did not predict these results and are contrasted with the present study with reference to simultaneous vs. sequential presentation of conflicting stimulus information.

VARIABLES ALTERING PERCEPTION OF THE ROTATING TRAPEZOIDAL ILLUSION

By CLARKE A. BURNHAM, University of Texas, and HIROSHI ONO, York University

Ames created the rotating trapezoidal-window illusion to demonstrate the effect of assumptions on perception.[1] Assumptions, according to the transactionalists, develop from action in the world:[2] because of our experience with the trapezoidal projection of rectangular distal objects, we assume and therefore see the rotating trapezoid as an oscillating rectangle. This interpretation of the role of assumptions was tested with the trapezoidal window by Allport and Pettigrew using a culture (the Zulu) in which the assumption of rectangularity was expected to be weak.[3] The results of their experiment have been interpreted by Slack as supporting a strong empiricistic position.[4] A second transactionalist proposition, *i.e.* that assumptions are modified when there is a discrepancy between a perception and an action based on that perception, has apparently not been tested with the trapezoidal window, however. Kilpatrick and Weiner have both found that the distorted-room illusion disappears with exposure to conflicting information derived from action, although this information apparently does not have to result from the viewer's own actions.[5] Experiments I and II of the present study were designed to clarify the role of action in the modification of the rotating trapezoidal illusion.

There have been studies demonstrating that the trapezoidal-

* Received for publication June 25, 1968. The research reported was completed while both authors were at Stanford University. The authors would like to thank Dr. Albert Hastorf for his advice.

[1] Adelbert Ames, Jr., Visual perception and the rotating trapezoidal window, *Psychol. Monogr.*, 65 (7), 1951 (No. 324).

[2] W. H. Ittelson, Perception and transactional psychology, in Sigmund Koch (ed.), *Psychology: The Study of a Science*, 4, 1962, 660–704.

[3] G. W. Allport and T. F. Pettigrew, Cultural influence on the perception of movement: The trapezoidal illusion among Zulus, *J. abnorm. soc. Psychol.*, 55, 1957, 104–113.

[4] C. W. Slack, Critique on the interpretation of cultural differences in the Ames trapezoid, this JOURNAL, 72, 1959, 127–131.

[5] F. P. Kilpatrick, Two processes in perceptual learning, *J. exp. Psychol.*, 47, 1954, 362–370; Melvin Weiner, Perceptual development in a distorted room: A phenomenological study, *Psychol. Monogr.*, 70 (16), 1956 (No. 423).

window illusion can be increased by verbal suggestion[6] and decreased by information about the illusion.[7] These results appear to contradict the transactional theory, since the experimental manipulations did not involve actions which provided information conflicting with the viewer's perception. Haber reported a sizeable decrease in oscillation after a period in which the illusion was discussed with *Ss*, but the transactionalists tend to deny that such cognitive information affects assumptions because the illusion persists for them, and for other investigators, even when they have such information. In the present study, Experiment III's use of Directed Viewers was designed to explore a possible mechanism by which both information and action can alter the rotating-window illusion.

There have also been studies of the cause of the apparent oscillation of the trapezoidal window. Although their accounts differ in details, Ames, Pastore, and Graham agree that a discrepancy between the perceived slant of the figure and its objective slant is necessary for the illusion of apparent oscillation.[8] As Epstein points out,[9] this explanation differs from that of Day and Power, who argue (from experiments with trapezoids and with other figures which also appear to oscillate when rotating) that the necessary condition for oscillation is an identity of the retinal projection of an object when rotating clockwise and counterclockwise.[10] Although the present study was primarily concerned with variables altering the illusion of oscillation, the cause of oscillation was investigated by studying the association between the perceived movement and the perceived shape of the rotating figure.

METHOD

Apparatus. The trapezoid used in the experiments was constructed from 3/64-in. sheet metal and painted flat gray. The long side of the figure was 13 in., the short side 9 in., and the horizontal length was 13⅜ in. A strip of ¾-in. masking tape was placed on the borders of the trapezoid to enhance its visi-

[6] J. M. McGee, The effect of group verbal suggestion and age on the perception of the Ames trapezoidal illusion, *J. Psychol.*, 56, 1963, 447–453; M. K. Cappone, The effect of verbal suggestion on the reversal rate of the Ames trapezoidal illusion, *J. Psychol.*, 62, 1966, 211–219.

[7] R. N. Haber, Limited modification of the trapezoidal illusion with experience, this JOURNAL, 78, 1965, 651–655.

[8] Ames, *loc. cit.*; Nicholas Pastore, Some remarks on the Ames oscillatory effect, *Psychol. Rev.*, 59, 1952, 319–323; C. H. Graham, On some aspects of real and apparent visual movement, *J. opt. Soc. Amer.*, 53, 1963, 1019–1025.

[9] William Epstein, *Varieties of Perceptual Learning*, 1967, 56–60.

[10] R. H. Day and R. P. Power, Apparent reversal (oscillation) of rotary motion in depth: An investigation and a general theory, *Psychol. Rev.*, 72, 1965, 117–127.

bility. A vertical spike ⅞ in. tall with a band of masking tape around it, making the spike ⅛ in. in diameter, was welded to the top of the long side. This spike provided a means of interaction with the rotating figure in Experiment I and a reference point for Ss' descriptions of perceived movement and shape. The trapezoid was rotated counterclockwise at a rate of 4.2 rpm. It was illuminated by two red 7½-w. bulbs placed below and to the side of the figure. Shields prevented Ss from seeing the bulbs. The figure was illuminated and visible only when it was rotating. The Ss viewed the trapezoid while seated in front of a table placed within a large reduction tunnel lined with black cloth. Ss viewed the trapezoid monocularly, using their preferred eye, from a distance of approximately 23 ft. The rear of the tunnel was 2 ft. behind the trapezoid. Shadows from the figure were not visible on the rear or side walls of the tunnel.

Procedure. The introductory comments, the sequence of viewing periods, and the method for obtaining Ss' perceptual reports were similar for the three experiments. The Ss were told that E was interested in reports of their phenomenology or "what the object looks like" instead of their "intellectual knowledge of what it is and what is happening." This emphasis on the appearance of the object was repeated in the instructions on the report sheet, which was labelled Perceptual Appearances Questionnaire. The Ss were first shown the rotating figure for a 45-sec. demonstration period during which it was ascertained that they could see the spike on the top of the figure. They were then shown the questionnaire and instructed how to complete it after later viewing periods.

There were four 5-min. experimental viewing periods. The Ss merely looked at the figure during Periods 1 and 4. After Period 1 all Ss were told that the figure was actually rotating in a counterclockwise direction, and E's interest in the apparent movement was restated. The Ss were never told that the figure was a trapezoid. The apparatus required for the action in Experiments I and II was introduced prior to Period 2, and the Ss engaged in the activity appropriate to their experimental treatment while viewing the figure during Periods 2 and 3. The apparatus was removed prior to Period 4. The Ss completed a questionnaire after each viewing period. They were asked to draw the pattern or patterns of movement they had seen during the preceding viewing period, to indicate the percentage of time they saw each pattern, and to draw the shape of the figure as it appeared for each pattern. The Ss were told that the apparent shape might change with changes in the pattern of movement. The patterns were drawn by indicating the movement of the end of the figure to which the spike was attached. Four sample patterns and the drawings to indicate them were presented on the front sheet of the questionnaire. These were counterclockwise rotation, clockwise rotation, oscillation with the spike end near, and a complex pattern of partial oscillation and rotation with a jump of the spike end from a twelve-o'clock position to a six-o'clock position. Pretesting had shown that these patterns were the most frequently observed.

Subjects. The Ss were 79 freshman and sophomore men at Stanford University. Approximately half were introductory-psychology students fulfilling a part of the course requirement; the others were recruited from a dormitory and paid for participation. The psychology students and the paid students were used in each of the experimental treatment groups.

Experiment I: String Actors and String Observers. The question asked in Experiment I was whether engaging in motor activity which conformed to the actual movement of the trapezoid would increase veridical perception. The *Ss* were run in pairs. They were seated beside each other with a partition between them. The Actor member of the pair held a stylus which was connected with a string to the spike of the trapezoid. He was instructed to follow the actual counterclockwise motion of the figure in a circular groove on the table while he kept the string taut. This was initially a difficult task. There was a considerable amount of strain on the trapezoid and motor when the Actor saw a reversal of direction and tried to pull the string toward him while the trapezoid was moving away. The Observer was given an unconnected stylus and told to trace the apparent movement of the trapezoid in the circular groove in front of him. Sixteen pairs of *Ss* were run; the Actor and Observer of each pair were selected to equate these conditions for mean percentage of veridical perception during Viewing Period 1. The data from one pair of *Ss* were excluded because the Actor stated after the experiment that he did not believe the string was attached to the trapezoid.

Experiment II: Weight Commanders and Weight Droppers. The question asked in Experiment II was whether instructing another person to engage in motor activity which provided information about the actual movement of the trapezoid would increase veridical perception. As before, *Ss* were run in pairs, seated beside each other with a partition between them. A stylus attached to a string was placed in front of one of the *Ss*. The string ran through a short vertical pipe placed over the trapezoid and had a weight at its end. The weight was visible in its raised position during the experimental viewing periods. When the stylus was pushed forward and the string was thus slackened, the weight dropped into position to be hit by the top edge of the rotating trapezoid when a side was at a half-past-four-o'clock position. This position was chosen because it was slightly in front of the point at which counterclockwise movement stopped during perceived oscillation of the trapezoid. One member of each pair of *Ss* was selected to be Weight Dropper; the other was Weight Commander. The Dropper's task was simple. He lowered the weight by moving the stylus forward in a groove in front of him whenever the Commander said "Drop!" The Dropper was instructed to lower and raise the weight rapidly and to keep looking at the moving figure. The Commander was told to order a drop whenever he thought there was a chance that the weight would hit the top of the figure near the edge with the spike. Commanders initially tended to order two drops during each rotation, one when the edge with the spike passed under the weight and the other immediately afterwards, when the edge appeared to return and again pass under the weight. This second drop provided visual information that the edge had continued rotating beyond the position of the weight. Sixteen pairs of *Ss* were run. The selection of Dropper and Commander was made to equate these conditions for mean veridical perception during the initial viewing period. The data from one pair were excluded because the Commander could not see the spike.

Experiment III: Directed Viewers. Reports from pretest *Ss* and from *Ss* in Experiments I and II suggested that an increase in veridical perception resulted from attention to certain portions of the rotating trapezoid. The question in Experiment III, then, was whether directing *S*'s attention to portions of the

figure would result in an increase in veridical perception. Following Period 1 S was told to try to see the actual counterclockwise rotation. He was given viewing instructions, based on the earlier reports, to help him. These instructions were (1) to focus on the corner with the spike rather than looking at the "object as a whole," (2) to look at one of the side edges of the figure, and (3) to look at the top edge and view this edge as a rod which was revolving around. The S was told to use any of these techniques, or others which he found useful, to help him see the rotation. Prior to Period 4 S was told to return to a relaxed or natural state of viewing while watching the object. Fifteen Ss were run individually in this experiment.

Results

Movement. Each S reported after each viewing period the movements he had seen during the period and the percentage of time he had seen each movement. All the reported patterns of movement except the veridical pattern of counterclockwise rotation were classified as illusory. Fig. 1 presents the mean percentage of veridical perception across the four periods for the 15 Ss in each of the five experimental treatments. The results of statistical analyses of the percentage of veridical perception (VP%) are as follows.[11] The Ss in the five experimental treatments did not differ in VP% in Period 1. There was a significant increase in VP% from Period 1 to Period 2 for all treatment groups except the String Observers. The VP% during Period 3 was significantly higher than during Period 1

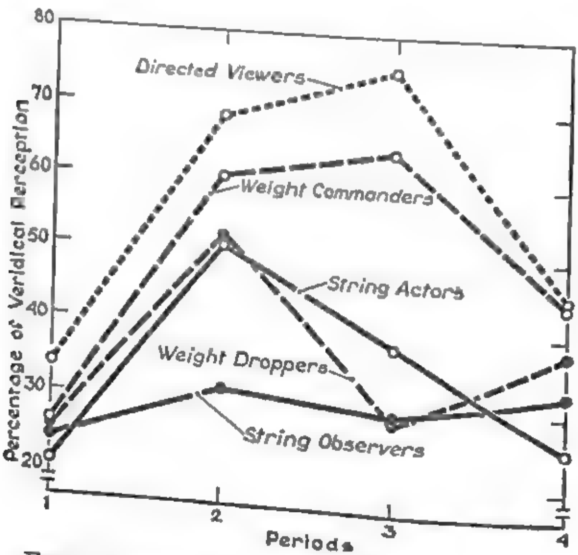


FIG. 1. PERCENTAGE OF VERIDICAL PERCEPTION

[11] All comparisons are based on two-tailed t tests with $\alpha = .05$.

for all except the String Observers and the Weight Droppers. The only significant change from Period 2 to Period 3 was the decrease of the Weight Droppers. Both the Weight Commanders and the Directed Viewers decreased significantly in VP% from Period 3 to Period 4. The only group with a significantly higher VP% during Period 4 than during Period 1 was that of the Weight Commanders. However, their final VP% did not differ significantly from that of any of the other four groups.

Movement and shape. In addition to reporting the movement of the figure, *S*s drew on the questionnaire the apparent shape of the figure for each pattern of movement they perceived. The drawn shapes were classified as rectangular if the average of the two interior angles on the longer side of the drawn figure was 86° or greater and as trapezoidal if the average of these angles was less than 86°.

The association between movement and shape was tested in two ways. For the first analysis, only data from those *S*s who reported seeing either counterclockwise rotation or oscillation for more than half of a viewing period were used. Data from *S*s who frequently saw other movement patterns were thus omitted from this analysis. The contingency classification of movement (counterclockwise rotation vs. oscillation) and shape (trapezoidal vs. rectangular) was analyzed by χ^2. The χ^2 values for the four periods and the usable *N* for each period are presented in Table I. Although these tests are not independent, it is quite clear that *S*s who tended to see the rotation saw a trapezoid while doing so and that *S*s who tended to see oscillation saw a rectangle while doing so.

The second method for studying the association between movement and shape answered the question whether an *S* who sees both counterclockwise rotation and oscillation will so report a change in his perception of shape with change in his perception of movement that there are more changes from rectangle to trapezoid than changes

TABLE I

ANALYSIS OF CONTINGENCY CLASSIFICATION OF MOVEMENT AND SHAPE

Period	Usable N	χ^2
1	59	8.56*
2	64	8.67*
3	59	9.24*
4	52	1.15

* $p < .01$, $df = 1$.

from trapezoid to rectangle associated with the change from oscillation to the rotation. The McNemar test for the significance of changes was applied to those *Ss* who saw both the rotation and oscillation in a viewing period. The χ^2 values for the four periods and the usable *N* for each period for this analysis are presented in Table II. These data also support the hypothesis that the perception of oscillation is associated with seeing a trapezoid and the perception of counterclockwise rotation with seeing a rectangle.

Discussion

What variables alter the perception of the rotating trapezoidal window and what mechanism causes the change in perception? Merely telling *Ss* that the object is rotating counterclockwise does not result in an increase in veridical perception. The String Observers, who were given this information, did not show an increase in veridical perception A simple interpretation of the transactionalists' statements on the importance of action is supported only by the comparison between the String Actors and the String Observers for Period 2, in which the action of the Actors led to an increase in veridical perception whereas observing did not. The 'hitches' encountered while acting might be said to have modified the *Ss'* assumptions. But the fact that the increase in veridical perception for the String Actors did not continue in Period 3, during which the action continued, or in Period 4 leads one to question the transactionalists' theorizing about the role of action on perception. The alteration in the perception of the Weight Commanders might be viewed as supporting a modified action hypothesis, similar to that expressed by Kilpatrick to account for his finding that both actors and nonactors changed their perception of the distorted room.[12] According to such a modified hypothesis, exposure to information conflicting with assumptions changes assumptions; this informa-

TABLE II

Analysis of Changes in Shape Accompanying Changes in Movement

Period	Usable *N*	χ^2
1	38	.90
2	33	7.11*
3	36	10.08*
4	41	7.11*

* $p < .01$, $df = 1$.

[12] Kilpatrick, *op. cit.*, 369.

tion is most frequently produced by one's own actions although it can be produced by another's. The Weight Commanders were certainly exposed to information which could change their assumptions. They did report a change in perception: they saw more counterclockwise rotation during Periods 2 and 3 and, to a lesser extent, Period 4 than during Period 1. However, the Weight Droppers were exposed to similar information, which resulted in an increase in veridical perception for Period 2 but not for Periods 3 and 4. Exposure to information resulting from actions does not necessarily alter the perception of motion of the rotating trapezoidal window. Indeed, the largest increase in veridical perception during Periods 2 and 3 occurred for the Directed Viewers. These *Ss* engaged in no gross motor activity and received no information from another's actions that the figure was rotating counterclockwise; they were, however, instructed to focus on a portion of the figure. Their data suggest that an increase in veridical perception is caused by selective focusing.

The mechanism of selective focusing is capable of explaining, in an admittedly a posteriori fashion, the pattern of alteration in motion perception for the groups in Experiments I and II. The task of the String Actors was initially difficult; selective focusing and, thereby, seeing rotation was functionally useful during Period 2. As these *Ss* continued to move the stylus around the grooved circle in front of them, the task of keeping the string taut became easier and could be performed even while they were perceiving oscillation. They probably did not need to selectively focus and see counterclockwise rotation during Period 3 to keep the string taut. The Weight Commanders had to order a drop every 4.2 sec. to perform their task correctly; seeing rotation would have been necessary for doing this during both Periods 2 and 3. Moreover, it is likely that they were led to fixate the corner of the trapezoid above the longer edge by the instructions to hit the figure at that edge where the spike was located. The Weight Droppers may have attempted to verify their Commanders' orders during Period 2, when the Commanders tended to order two drops during each rotation. Thus they may have engaged in selective focusing during this period in which their perception changed. They probably performed their task more mechanically during Period 3 and did not attempt to see counterclockwise rotation.

In sum, we suggest that the action of the *Ss* in Experiments I and II resulted in selective focusing when it was necessary or useful to

perceive the veridical movement pattern to support the action. It is unlikely that a change in assumptions or a change in the weighting of assumptions brought to the experimental situation was responsible for the alterations in perceived movement. The lack of continued veridical perception of motion during Period 4 is incompatible with an hypothesis that assumptions were modified. When one focuses on one edge of the figure, one sees rotation; when one looks at the figure in a relaxed fashion, one sees oscillation. This mechanism is also capable of explaining previous findings of changes in the perception of the rotating trapezoidal-window illusion. It is possible that the experimental demand characteristics of McGee's and Cappone's instructions resulted in more relaxed viewing and thus in an increase in oscillation and that Haber's information sessions caused his *Ss* to see rotation by selective focusing.[13] Selective focusing can also account for the relative stability of the illusion for professionals who have worked with it and who have not been motivated to destroy it. Both *Es* of the present study became adept at alternating between perceiving counterclockwise rotation and oscillation by alternating between selective focusing and relaxed viewing.

Selective focusing may result in veridical perception for two reasons. An *S* who fixates an edge may be able to disregard the object-connotation aspects of the rotating trapezoidal window and therefore see the figure as a rotating trapezoid. This interpretation of the effect of selective focusing is based on an assumption explanation of the illusion such as that proposed by Ames.[14] Or, fixating one edge may result in a failure to notice the illusory perspective cues which can produce apparent oscillation. This interpretation is based on a linear-perspective explanation of the illusion such as that proposed by Graham.[15] The present study does not allow one to differentiate between these two possible explanations of the effect of selective focusing. Analysis showed that one shape, a rectangle, was associated with one perceived movement, oscillation, and that another shape, a trapezoid, was associated with the perception of counterclockwise rotation. These results tend to support a slant-shape interpretation of the rotating trapezoidal illusion rather than an ambiguity-of-retinal-motion interpretation. Day and Power's interpretation of oscillation in terms of an equality of projected motion for both directions of rotation may be reasonable for a figure

[13] McGee, *loc. cit.*; Cappone, *loc. cit.*; Haber, *loc. cit.*
[14] Ames, *op. cit.*, 13–17.
[15] Graham, *loc. cit.*

such as an ellipse.[16] There is no provision within their theory for a change in shape accompanying a change in perceived movement as there is within the misperception-of-slant interpretation of oscillation. It is possible that shape and movement data could be obtained which would support one of the two interpretations of the effect of selective focusing. The interpretation that focusing prevents the assumption of rectangularity from occurring implies that the change in shape precedes the change in movement. The interpretation that focusing prevents the illusory perspective cues from being registered implies that the change in movement precedes the change in shape. However, our shape and movement data are correlational, and thus we cannot tell which variable is causal or whether they are concomitant.

Summary

Three experiments with the rotating trapezoidal-window illusion were conducted which bear upon the transactionalists' proposition that perceptions are modified by action. The results suggest that action can lead to an increase in veridical perception but that it does so through the mechanism of selective focusing. *S*s told to fixate an edge of the figure reported a greater increase in veridical perception than *S*s who engaged in gross motor activity with the figure. It is hypothesized that action results in selective focusing which produces veridical perception when such perception is necessary to support the action.

[16] Day and Power, *loc. cit.*

ASSOCIATIVE RECALL AS A FUNCTION OF STIMULUS RECOGNITION

By JAMES M. ROYER, University of Illinois, Urbana

Martin has recently reported two experiments concerned with the role of stimulus recognition in paired-associate learning.[1] In these experiments the position was taken—and the results generally confirmed—that in order for a stimulus to elicit its appropriate recall response, *S* must first make an intervening recognition response to that stimulus. The purpose of the present paper was to shed some light on an anomaly found in the second of Martin's studies, where he used a mixed-list, study-trial/test-trial design and two types of CCCs of equal associability.[2] Six CCCs of the first type (Type C) were consistently paired with one of the digits 1–3 throughout the 12 study trials, while six CCCs of the second type (Type R) were re-paired anew on each study trial. The test trials consisted of presentation of the 12 stimuli (Types R and C) from the study trials and 12 additional CCCs of equal associability that had not appeared on previous study or test trials. Martin's *Ss* were then required to make both a recognition response and a recall response to each test stimulus, but the results of the study indicated less support for Martin's hypothesis than might have been desired. The data indicated a rather systematic improvement over trials in the proportion of recall responses correctly associated with nonrecognized stimuli for both R and C pairs. In addition, the proportion of correct associations made to nonrecognized R stimuli significantly exceeded the chance level. Martin attributed the nonchance proportion of correct recall responses given the nonrecognized Type R stimuli to sampling error. He has written (personal communication) that this suggestion was truly residual. Repeated analyses of the data failed to turn up any meaningful alternative explanations.

* Received for publication August 12, 1968.

[1] Edwin Martin, Stimulus recognition in aural paired-associate learning, *J. verb. Learn. verb. Behav.*, 6, 1967, 272–276; Edwin Martin, Relation between stimulus recognition and paired-associate learning, *J. exp. Psychol.*, 74, 1967, 500–505.

[2] Martin, *op. cit.*, *J. exp. Psychol.*, 74, 1967, 500–505.

However, the systematic improvement over trials for proportion of correct recall responses given nonrecognition of the stimulus for both R and C pairs has a possible explanation. There is the possibility that such a systematic increase over trials could have occurred if at least some of the *Ss* had adopted a particular guessing strategy while going through the study-and-test trials. The response items in Martin's study, for both C and R pairs, were the digits 1, 2, or 3, and each number was used four times within each study-and-test trial, *i.e.* paired twice with a C item and twice with an R item. Now let us suppose that an *S* on the first test trial had recognized and correctly recalled the first two items, both of which had the digit 1 as the correct response. If on the third item *S* was required to recall a response for a stimulus that he did not recognize, he could improve his probability of correct recall over chance by guessing the digit 2 or 3 as a response for that item. Assuming that some of the *Ss* did use such a strategy, then as they proceeded through the study-and-test trials and learned more of the correct S-R associations, they would be able to increasingly improve their guessing performance by elimination of one or more of the alternative responses on items for which there was not certain recognition. Thus, one way to control for such an occurrence would be simply to make it more difficult for *Ss* to guess the appropriate response by increasing the number of responses available to them. The present study accordingly attempted to replicate Martin's second study with the number of available responses increased from three to eight. The increase also allowed a more random assignment of response digits in that a single digit did not have to appear a specified number of times on any single trial. Under this procedure it was expected that *Ss* would demonstrate that stimulus recognition is necessary in order to make a correct associative response.

Method

Subjects, design, and materials. Ss were 24 college students (men and women) who volunteered as paid participants. Twelve students were assigned to each of the two paired-associate lists and were given 12 study-and-test trials on an aurally presented list of 12 trigram-number pairs. The stimuli were CCCs of 45%–55% association value,[3] and the responses were the digits 1–8. The 12 CCCs on each list were divided into two groups of six each. Those in the first group of trigrams (Type C) were consistently paired with the same digit throughout the 12 study trials. Those in the second group of trigrams (Type

[3] L. Witmer, The association value of three-place consonant syllables, *J. genet. Psychol.*, 47, 337–360.

R) were re-paired anew at the end of each study trial. The re-pairing of Type R stimuli with new digits was conducted in a random fashion except for the restrictions (1) that no digit be used as a response more than three times on a single trial and (2) that there be an interval of seven trials before the same digit was re-paired with the same CCC. Two 12-pair lists were constructed so that every trigram entered into a Type C pair on one list and into a Type R pair on the other list. Twelve *Ss* were assigned to each list.

During test trials 24 trigrams were presented: the 12 Type C and Type R trigrams from the study trials and 12 'filler' trigrams of equal association value. The filler trigrams were presented only once each during the 12 test trials. For each of the trigrams presented on a test trial, *S* was required to make two responses. He first had to make a recognition or a nonrecognition response and then to make a recall response, *i.e.* one of the digits, 1-8. The 12 test stimuli and the 12 filler trigrams were intermixed in a different random order on each test trial.

Materials for the study were prerecorded on tape and presented to groups of six *Ss* in a small room. *Ss* recorded their responses in a 13-page test booklet. On each page were 24 blank spaces for recording recall responses and beside each blank were the letters N and O for recording recognition responses. *Ss* were required to place a slash through the letter O ("old") if they recognized the stimulus as one that appeared on a previous study trial or through the letter N ("new") if they judged it to be a new stimulus.

Procedure. Each group of six *Ss* was told that the task consisted of a series of 12 study-and-test trials. *Ss* were instructed to listen to the series of 12 trigram-number pairs during the study trial and told that the pairs were to be presented at a 2-sec. rate. They were also told that each study trial would be followed immediately by a test trial in which they would hear the trigrams alone and that they were to record the appropriate responses in the test booklet. *Ss* were further advised that some of the CCCs would change responses from trial to trial.

Following reading of the instructions a sample study-and-test trial was run; this consisted of CCCs which appeared only on that trial. *Ss* recorded their responses on the first page in the booklet. The tape was then stopped and the *Ss* were asked if they had any questions. The tape was then started again with the first study trial. Study-trial items were presented at a 2-sec. rate and the test-trial items at a 6-sec. rate. The intertrial interval was 5 sec., during which time *Ss* were told whether the next trial was a study trial or a test trial.

Results

Recognition. A correct recognition occurred when *S* placed a slash through the O on the test booklet following a test-trial trigram that had been presented as a study-trial trigram. The proportion of recognition over trials is presented in Fig. 1. Collapsing over trials, the proportions of recognition were .73 for Type C pairs and .706 for Type R pairs. There was no statistical difference between recognition of Type C and Type R pairs ($Z = 1.57$, $p = .12$).[4] The

[4] Standard *Z* test for the difference between two proportions; see H. Walker and J. Lev, *Statistical Inference*, 1953, 77.

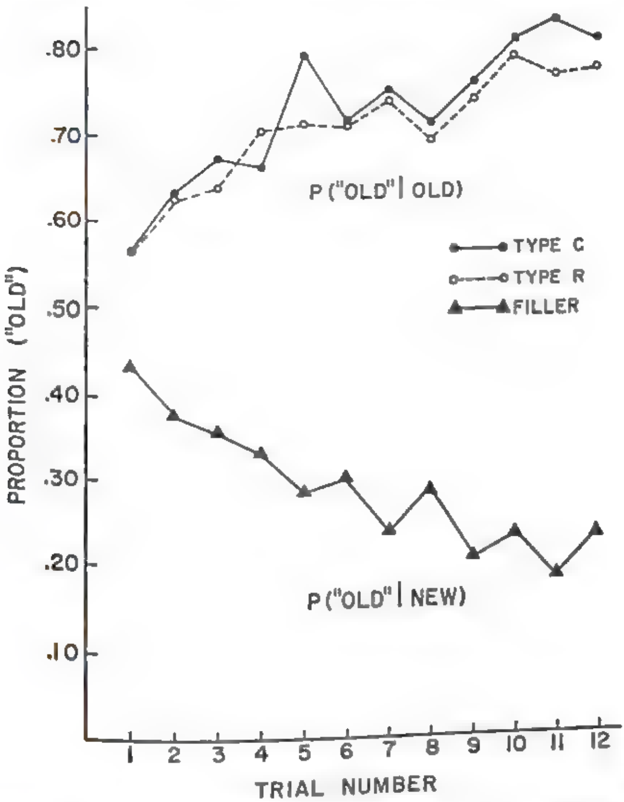


FIG. 1. PROPORTION "OLD" RESPONSES TO "OLD" (STUDY-TRIAL STIMULI, TYPES C AND R) AND "NEW" (FILLER) TRIGRAMS

proportion of false recognition, *i.e.* indicating "old" for a filler trigram, is shown in the bottom half of Fig. 1.

Learning. The proportion of correct recall responses to study-trial stimuli as a function of test-trial number is shown in Fig. 2. As can be seen, the proportion of correct recall for Type C pairs given recognition of the stimulus (P[CR/R]) reached a higher level than did the same proportion for Type R pairs. The differences in proportions of correct recall given recognition (P[CR/R]) for Type C and Type R pairs were compared over blocks of four trials, *i.e.* over Trials 1–4, 5–8, and 9–12. The differences were, respectively, $Z = .207$ ($p = .79$), $Z = 3.09$ ($p = .002$), and $Z = 6.35$ ($p < .001$). Collapsed over trials, the statistic was $Z = 7.0$ ($p < .001$), making an initially nonsignificant comparison highly significant.

Learning-recognition contingency. Table I summarizes the information on the learning-recognition contingency. Note that the prob-

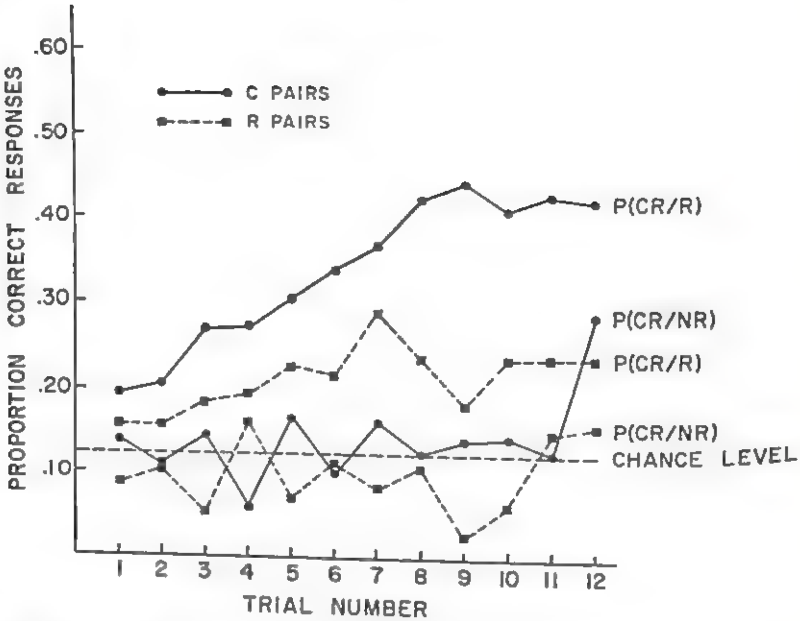


Fig. 2. Proportion Correct Recall (CR) to Study-Trial Stimuli for Type C and R Pairs Given Recognition (R) and Nonrecognition (NR)

ability of a correct recall response by chance alone was set at .125, since there were eight alternatives. The proportion of correct recall for Type C pairs given recognition of the stimulus (P[CR/R]) was .345. Compared to the chance level, .125, this yielded a highly significant statistic ($Z = 13.09$, $p < .001$). In contrast, the proportion of correct recall given nonrecognition of the stimulus (P[CR/NR]) was .139, which, when compared to chance, yielded

TABLE I

Proportion Correct Recall (CR) Given Recognition (R) and Nonrecognition (NR) of the Stimulus

	Pair type	
	C	R
P(CR/R)	.345	.219
P(CR/NR)	.139	.100
P(chance CR)	.125	.125
P(CR/R) on trial blocks		
1–4	.237	.175
5–8	.356	.254
9–12	.420	.223
P(CR/NR) on trial blocks		
1–4	.119	.103
5–8	.138	.098
9–12	.176	.097

a nonsignificant statistic ($Z = .636$, $p = .535$). Further, the proportion of correct recall given recognition significantly exceeded the proportion of correct recall given nonrecognition ($Z = 8.44$, $p < .001$).

Comparisons for the Type R pairs were made in the same manner. The proportion of correct recall for the Type R pairs given recognition of the stimulus (P[CR/R]) was .219. This value, when compared to chance level, indicated a statistically significant above-chance performance ($Z = 6.26$, $p < .001$). The same comparison, made with P(CR/NR) = .10, indicated no significant departure from chance level ($Z = -1.26$, $p = .173$). The comparison of P(CR/R) with P(CR/NR) for Type R pairs revealed significantly higher performance given recognition of the stimulus ($Z = 5.83$, $p < .001$).

In comparisons between Type C and Type R pairs, as shown above, P(CR/R) for Type C pairs significantly exceeded the same statistic for Type R pairs ($Z = 7.0$, $p < .001$). In addition, the comparison of P(CR/NR) for the two pair types yielded no differences at the .05 level ($Z = 1.88$, $p = .0602$). The proportion of correct recall, given recognition and nonrecognition of the stimulus, is further broken down in Table I to give blocks-of-trials comparisons. Of special interest here is the P(CR/NR) for Type R and Type C pairs. None of these proportions departed significantly from chance. The largest departure from chance, that for Type C pairs on Trials 9–12, yielded $Z = 1.06$, $p = .289$.

Discussion

The findings of the present study were essentially in agreement with the second Martin study. The present study found, as did Martin's, that there was no difference between Type R and Type C pairs in an *S*'s ability to recognize them as having occurred on a previous study trial. However, there was a significant difference in favor of Type C pairs when *S* was asked to recall the response associated with the trigram. As has been indicated, this finding suggests that formation of an associative (recall) response is not a requirement for formation of a recognition response in a paired-associate learning situation.[5]

The major area of departure between the present study and Martin's study lies in the likelihood of a correct recall response

[5] Martin, *op. cit.*, *J. exp. Psychol.*, 74, 1967, 500–505.

given nonrecognition of the stimulus. Martin was forced to make no decision concerning his stronger hypothesis, *i.e.* that formation of a recognition response was prerequisite to formation of an associative response. It was suggested in the present study that the results of Martin's study could have occurred if some of the *Ss* were using a guessing strategy allowing them to perform above chance level on nonrecognized items. It was further suggested that with the addition of several more alternative responses such a guessing strategy would become unlikely. Four specific comparisons can be made between the present study and Martin's study to clarify this area of departure. (1) Martin's *Ss* achieved significantly above-chance correct recall given nonrecognition of the stimulus for Type R pairs ($Z = 2.0$, $p = .05$). The comparable statistic for the present study was nonsignificant ($Z = -1.26$, $p = .173$). (2) Martin's data indicated a significant difference in P(CR/NR) for R and C pairs ($Z = 2.28$, $p = .02$). The difference in the present study was large, though nonsignificant at the .05 level ($Z = 1.88$, $p = .0602$). (3) Two of the possible six comparisons of P(CR/NR) against chance over blocks of four trials for both R and C pairs proved significant in the earlier study. None of the comparisons in the present study proved significant. And (4) there was an apparent improvement over blocks of trials in P(CR/NR) for both Type R and Type C pairs in the Martin study. Table I reveals that the comparable statistics for the present study were .119, .138, and .176 for the Type C pairs and .103, .098, and .097 for the Type R pairs. It can be seen that there was an apparent improvement over trials for the Type C pairs; however, no such trend existed for Type R pairs.

Thus, the present study cannot offer conclusive evidence that *Ss* in the previous study engaged in a guessing strategy that enabled them to improve over trials in their recall of unrecognized items. It does, however, offer substantial support to Martin's contention that an intervening recognition response must occur before a particular stimulus elicits an associated (recall) response. *Ss* performed consistently at the chance level given nonrecognition of the stimulus and consistently above chance given recognition of the stimulus.

Summary

The study-trial/test-trial method was used with two types of aurally-presented paired associates. The paired associates consisted of 12 trigram-number pairs, six of which were consistently paired throughout the 12 trials, and six of which were re-paired anew at

the end of each study trial. During test trials *S*s were required to make a recognition response and a recall response to both the trigrams presented on the study trials and to 12 new filler trigrams. It was suggested that recall of the correct response item would be dependent on *S*'s ability to recognize the stimulus. The data supported this hypothesis.

EFFECTS OF READY-SIGNAL INTENSITY AND INTENSITY OF THE PRECEDING RESPONSE SIGNAL ON SIMPLE REACTION TIME

By David L. Kohfeld, Psychology Division, U. S. Army Medical Research Laboratory, Fort Knox, Kentucky

Recent research has shown that stimulus-intensity effects in simple reaction time (RT) depend not only upon the absolute stimulus values but also upon their difference from some subjective reference level. At least two experimental approaches have proved useful in clarifying these relationships. (1) Grice and Hunter demonstrated that within-*S* designs produce larger differences in RT to varying signal intensities than when each response signal is presented to a separate group.[1] They suggested that a reasonable interpretation of the contrast effect found in within-*S* designs is provided by adaptation-level (AL) theory.[2] (2) It has been demonstrated that when *S*s listened to brief presentations of a tone just prior to participation in an RT experiment, the resulting response-signal functions were modified in accordance with the intensity of the preadaptation stimulus.[3] Clearly, the stimulus context in which a RT signal is presented is a significant factor in determining the speed of response.

An important feature of simple RT experiments is that a ready-signal is usually presented on each trial to facilitate the preparation for reaction, and since the ready-signal precedes the response signal on each trial, it seems reasonable that the intensity of the former may have important consequences for the manner in which the latter is perceived. In support of this contention, Baumeister,

* Received for publication August 12, 1968.

[1] G. R. Grice and J. J. Hunter, Stimulus intensity effects depend upon the type of experimental design, *Psychol. Rev.*, 71, 1964, 246–256.

[2] Harry Helson, *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior*, 1964. Grice and Hunter proposed that exposure to several intensities of test stimuli produces an AL intermediate among the values, whereas exposure to only one intensity results in an AL at that particular value.

[3] H. G. Murray and D. L. Kohfeld, Role of adaptation level in stimulus intensity dynamism, *Psychon. Sci.*, 3, 1965, 439–440; D. L. Kohfeld, Stimulus intensity and adaptation level as determinants of simple reaction time, *J. exp. Psychol.*, 76, 1968, 468–473.

Dugas, and Erdberg, employing auditory ready- and response signals, found that increases in ready-signal intensity produced reliable increases in RT for all response-signal values.[4] They suggested that the ready-signal contributed to the effective stimulus context, but they did not relate their findings to a specific theoretical position. The present study was an attempt to determine whether the ready-signal can influence RT by serving as a reference stimulus with which the response signal is compared. Accordingly, it was hypothesized that variations in ready-signal intensity would produce corresponding changes in RT. Based on the findings of preadaptation studies, it was expected that the highest ready-signal intensities would provide the slowest mean RT and that lowest ready-signal intensities should produce the fastest mean RT.[5]

Method

Apparatus. The stimuli for both the ready- and response signals were 1000-~ tones ranging in intensity from 30 to 90 db., SPL. The tones were generated by a Krohn-Hite audio oscillator and presented through calibrated earphones. One Grason-Stadler electronic switch, having a rise and decay time of 100 msec., presented the ready-signal, while a second switch, with a rise and decay time of 10 msec., delivered the response signal. The rise and decay times were set at this level in order to avoid click transients. The experiment was conducted in a double-walled, sound-treated chamber. *S* sat in a chair with a conventional telegraph key clamped on its arm.

Subjects and procedure. The *Ss* were 52 soldiers (average age of 20 years) who were assigned to the laboratory after completion of basic training. Each *S* was given conventional RT instructions before beginning the experiment. It was emphasized that *S* should press the key as fast as he could to the second of two successive tones. A total of 96 scored trials per session was presented. Before the first session 15 unscored practice trials were given to familiarize *S* with responding to the second of two tones.

The ready-signals were 30-db., 60-db., or 90-db. tones of .5-sec. duration. Forty *Ss* were randomly assigned to four ready-signal conditions of either 30 db., 60 db., 90 db., or a random combination of the three (30–60–90-db.), 10 *Ss* in each condition. Each ready-signal was presented an equal number of times in each block of 24 trials for the 30–60–90-db. condition. The remaining 12 *Ss* were given all four ready-signal conditions in a counterbalanced order on consecutive days. All *Ss* were administered the same order of 30-, 50-, 70-, and 90-db. response signals, presented in random order with the restriction that there be six presentations of each response signal in each block of 24

[4] A. A. Baumeister, Jeanne Dugas, and Philip Erdberg, Effects of warning signal intensity, reaction signal intensity, preparatory interval, and temporal uncertainty on reaction times of mental defectives, *Psychol. Rec.*, 17, 1967, 503–507.

[5] Murray and Kohfeld, *op. cit.*, 440; Kohfeld, *op. cit.*, 473.

trials. Foreperiods of 1, 2, or 3 sec. were given in irregular order preceding the onset of the response signal, which lasted for 1.5 sec. The interval between response-signal offset and ready-signal onset was set at 15 sec. Timing of events was controlled by Hunter timers operating in a repetitive sequence. RT was recorded in msec. by a Hunter electronic timer.

Results

The left panel in Fig. 1 shows mean RT as a function of response-signal intensity for the four ready-signal conditions. An analysis of variance indicated that the within-*S* effect of response-signal intensity was highly significant, $F\ (3, 108) = 124.91$, $p < .001$. Also significant was the main effect of ready-signal intensity, $F\ (3, 36) = 3.22$, $p < .05$. It was apparent that mean RT was fastest for the 30-db. ready-signal, slowest for the 90-db. ready-signal, and intermediate for the 60-db. and 30–60–90-db. conditions. The fact that the 60-db. and 30–60–90-db. groups performed at intermediate levels suggested that both groups were responding to a value at the mean of the stimuli (60-db.). A test for simple effects revealed that these two groups did not differ significantly, $t\ (18) = 1.06$. In order to determine whether the

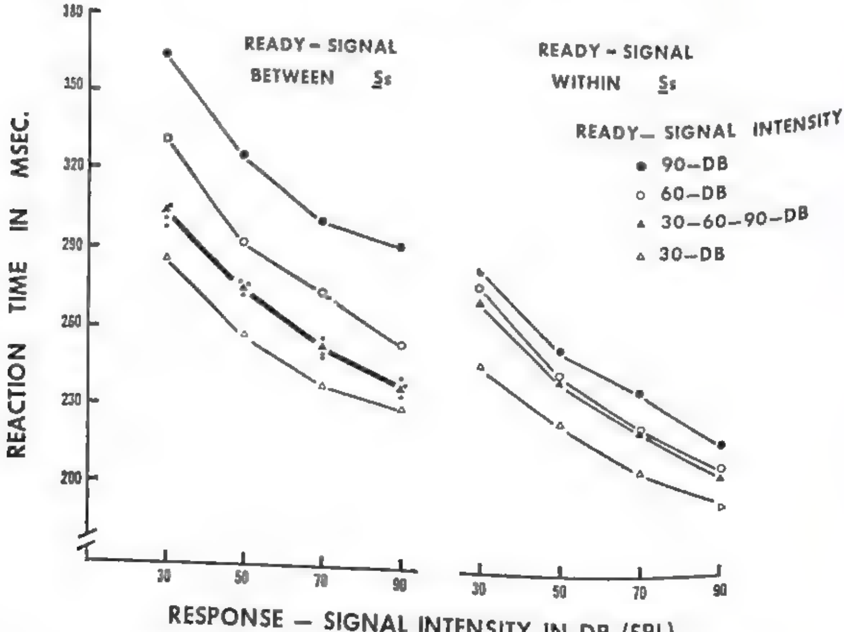


FIG. 1. REACTION TIME AS A FUNCTION OF READY- AND RESPONSE-SIGNAL INTENSITY

(The dots around the between-*S*s 30–60–90-db. ready-signal means are plots of the separate 30-db., 60-db., and 90-db. trial components.)

30–60–90-db. group responded differentially to the 30-, 60-, and 90-db. ready-signal trials, each ready-signal was plotted separately in the left panel of Fig. 1. The 30-, 60-, and 90-db. trial components showed small and unsystematic departures from the group mean. It appears that *S*s in this group responded consistently to a value at the mean of the ready-signals in spite of random trial-to-trial shifts in ready-signal intensity.

In view of the stability of ready-signal effects within an RT session, as evidenced by the trial-to-trial analysis of ready-signal components in the 30–60–90-db. group, the four ready-signal conditions were presented to 12 *S*s in a counterbalanced order to determine whether RT would change from day to day as a function of ready-signal intensity. The right panel of Fig. 1 reveals that RT changed over days, as *S*s were fastest on the 30-db. day, slowest on the 90-db. day, and intermediate on days when the 60-db. and 30–60–90-db. ready-signals were presented. An analysis of variance indicated that the within-*S* effects due to ready- and response-signal intensity were significant, $F(3, 33) = 3.50$, $p < .025$, and $F(3, 33) = 196.07$, $p < .001$, respectively. The 60-db. and 30–60–90-db. ready-signal conditions did not produce significant differences in RT, $t(11) = .39$, which confirms the prediction of similar response potential for the two conditions.

It has been suggested that trial-to-trial variations in response-signal intensity should displace *S*s' AL toward the intensity of the response signal on the preceding trial.[6] In order to evaluate this assumption, Fig. 2 presents RT to response-signal intensity on trial n as a function of the response-signal intensity on trial $n-1$.[7] The data points represent mean RT averaged over ready-signal conditions for the 12 *S*s in the within-*S* design. It is apparent from Fig. 2 that RT on trial n could not be predicted on the basis of trial $n-1$ response-signal intensity. Supporting this conclusion were the results of an analysis of variance which indicated that the trial $n-1$ effect was not significant, $F(3, 33) = .33$. Moreover, the absence of this effect was consistent across ready-signal conditions, as evidenced by the lack of a Ready-Signal × Trial $n-1$ interaction, $F(9, 99) = .81$. Clearly, trial-to-trial changes in response-signal intensity did not produce corresponding shifts in AL.

[6] I. D. John, Sequential effects of stimulus characteristics in a serial reaction time task, *Aust. J. Psychol.*, 19, 1967, 35–40.

[7] The author wishes to thank H. G. Murray for suggesting this analysis.

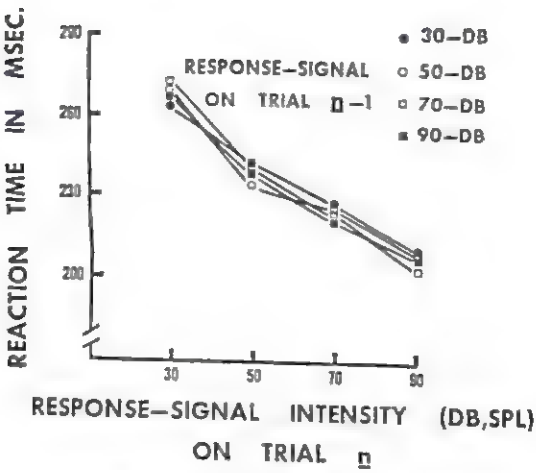


Fig. 2. Reaction Time to Response-Signal Intensities on Trial *n* as a Function of the Response-Signal Intensity Presented on the Previous Trial (*n*—1)

Discussion

The most noteworthy aspect of the present results was that the intensity of the ready-signal contributed to the effective stimulus context in which the response signals were presented. This finding is consistent with previous attempts to describe the departure of a stimulus from AL as an important determinant of response potential in the RT situation.[8] A feature of the present design which deserves comment is that both the ready- and response-signal intensities were apparently important in the formulation of a reference level. Thus, pronounced within-*S* intensity effects were obtained for the response signals, a finding which supports the notion that a contrast effect can be obtained when *S* experiences more than one response-signal value in an RT session. The intensity of the ready-signal also produced a contrast effect, a result which is consistent with the findings of Baumeister, Dugas, and Erdberg.[9] As to why the ready-signal produces this effect, it is argued that *S*s are required to attend closely to the ready-

[8] Grice and Hunter, *op. cit.*, 252; Murray and Kohfeld, *op. cit.*, 440–441; Kohfeld, *op. cit.*, 468–473.

[9] Baumeister, Dugas, and Erdberg, *op. cit.*, 505–506. It should be noted that Isaac Behar and C. K. Adams (Some properties of the reaction-time ready-signal, this Journal, 73, 1966, 419-426) reported that RT showed a systematic decrease with a corresponding increase in ready-signal intensity. This apparent contradiction in findings may be due to differences in methodology, as Behar and Adams employed a design more typical of classical conditioning (only one response signal on all trials; little variation in foreperiod interval) than of simple RT.

signal, thus enhancing its contribution to the weighting process which apparently occurs in the RT situation.

There were two aspects of the results which pointed to the conclusion that trial-to-trial changes in the stimuli did not induce compensatory adjustments in the reference level of the observer. (1) There were no fluctuations in RT as a function of the response-signal intensity presented on the preceding trial. (2) When the three ready-signal intensities were presented in irregular order within a session, *Ss* apparently responded to their mean value and did not show trial-to-trial adjustments toward the prevailing ready-signal value. A possible explanation for this stability is based on the assumption that *S* was unable to modify his reference level from trial to trial because the signal intensities were presented in an unpredictable order. As a consequence of this uncertainty, *S* chose an effective level which was intermediate among the intensity values.

The absence of response-signal sequence effects in the present study is somewhat surprising in view of the findings of both John and Murray, who reported that RT was an increasing function of the response-signal intensity presented on the previous trial.[10] Two possible reasons for the discrepant findings are offered. (1) The present study employed an auditory ready-signal, whereas Murray used a visual ready-signal, and John used none at all in his serial RT study. One might speculate that auditory response-signal sequence effects cannot be observed when a ready-signal of the same modality is presented between response signals. (2) The present study utilized an intertrial interval of 15 sec., whereas John employed an average interstimulus interval of 6.5 sec., and Murray used an intertrial interval of 6 sec. In discussing his results, Murray suggested that each response signal may leave a residual "neural noise" which influences the context in which the next signal is presented.[11] It follows that the residual effects of the response signal may persist over a relatively short intertrial interval (*e.g.* 6 sec.) but not over a relatively long intertrial interval (*e.g.* 15 sec.).

The fact that adaptation effects were so persistent deserves a word of further comment in the context of AL theory. Since the theory predicts that every stimulus produces an adjustment of

[10] John, *op. cit.*, 38; H. G. Murray, unpublished doctoral dissertation, University of Illinois, 1968. 46–48.
[11] Murray, *op. cit.*, 53.

the prevailing level, the inability of *Ss* to modify their AL in the direction of the stimuli presented on each trial is seemingly inconsistent with an AL interpretation of sensory phenomena. This issue can be partially resolved, perhaps, by making the rather obvious point that something more than sensory processes is reflected in RT measures. Indeed, Helson has suggested that AL is not a purely sensory phenomenon but is also manifested in certain cases where motor responses, such as RT, represent the prevailing AL.[12] Furthermore, Grice has pointed out that there is a qualitative difference between ALs which involve a modification of sensory processes and those which can more appropriately be classified as response processes.[13] Grice's main proposal was that adaptation procedures, along with a number of other variables, can be viewed as influencing the detection criterion, a concept which is derived from decision theory and which in many ways is analogous to the AL concept. This approach not only allows one to allocate certain adaptation effects within the general domain of response processes, but it also implies that adaptation procedures which influence the criterion can be quite persistent, an implication which is supported by the present experiment.

SUMMARY

Forty *Ss* (men) were given a simple RT test with four ready-signal conditions, 10 *Ss* in each condition. A fifth group of 12 *Ss* received all four ready-signal conditions in a counterbalanced order on consecutive days. The results indicated that a 90-db. ready-signal produced the slowest mean RT, 30-db. the fastest, 60-db. intermediate, while a random combination of ready-signals with a mean of 60 db. also produced intermediate RTs. It was concluded that the ready-signal can influence RT by serving as a reference stimulus with which the response signal is compared. An additional analysis revealed that there were no fluctuations in RT as a function of the response-signal intensity presented on the preceding trial, a finding which suggested that trial-to-trial changes in the stimuli did not induce compensatory adjustments in the effective AL of the observer.

[12] Harry Helson, personal communication, 1967.

[13] G. R. Grice, Stimulus intensity and response evocation, *Psychol. Rev.*, 75, 1968, 372.

MOTIVATION AND SHORT-TERM RETENTION: EVIDENCE FOR COVERT REHEARSAL

By Roger M. Tarpy, Williams College,
Sam Glucksberg, and Loy D. Lytle,
Princeton University

Facilitative effects of incentive upon short-term memory performance have been demonstrated in a number of recent studies.[1] The general experimental procedure, a variant of the Peterson and Peterson technique,[2] involves the presentation of a recall stimulus (*e.g.* a CCC trigram), followed by an interpolated rehearsal-preventing activity (*e.g.* reading digits in time to the beat of a metronome). After a specified time interval filled by this activity, *Ss* try to recall the stimulus. Differential incentives for recall may be cued in a number of ways. One example would be by varying the background color of either the recall stimuli or the interpolated-task (IT) stimuli, so that IT stimuli presented on a red background would, say, inform *S* that correct recall is worth 10¢, while IT stimuli on a blue background would indicate that correct recall is worth nothing. Although differential learning may occur under some conditions, incentive effects have been obtained when differential learning was precluded.[3]

Two alternative explanations for these incentive effects have been proposed, an arousal explanation and a covert-rehearsal explanation. Weiner suggests that "differences in recall between motivational conditions and nonmotivational conditions . . . [are] not caused by differential rehearsal . . . of stimuli."[4] Incentive

* Received for publication February 2, 1968. The research reported was supported in part by Public Health Service Grant MH-10742-03 to Princeton University, S. Glucksberg, principal investigator.

[1] T. Kernoff, B. Weiner, and M. Morrison, Affect and short-term retention, *Psychon. Sci.*, 4, 1966, 75–76; R. M. Tarpy and S. Glucksberg, Effects of incentive and incentive-cue position on short-term retention, *Psychon. Sci.*, 5, 1966, 313–314; B. Weiner and E. L. Walker, Motivational factors in short-term retention, *J. exp. Psychol.*, 71, 1966, 190–193.

[2] L. R. Peterson and M. Peterson, Short-term retention of individual verbal items, *J. exp. Psychol.*, 58, 1959, 193–198.

[3] Tarpy and Glucksberg, *loc. cit.*; B. Weiner, Motivation and memory, *Psychol. Monogr.*, 80, 1966, No. 18, Whole No. 626.

[4] Weiner, *loc. cit.*

effects are attributed to some as yet unspecified arousal function such that ". . . augmented motivation during trace formation makes . . . [that trace] more resistant to interference" or, presumably, more resistant to decay.[5] In support of this arousal interpretation, Weiner argues that if differential covert rehearsal does mediate incentive effects, then increasing the difficulty of the IT should decrease incentive effects because the greater the IT difficulty, the less the opportunity for covert rehearsal.[6] Weiner found no Incentive × IT Difficulty interaction, and he concluded that augmented recall is not mediated by covert rehearsal. This conclusion may be premature because it rests on the implicit assumption that *Ss'* attention to the IT increases in proportion to its difficulty. Weiner did not assess performance on the ostensibly rehearsal-preventing IT. It is conceivable that *Ss'* accuracy in the IT fluctuates as a function of incentive and difficulty and that differential covert rehearsal could, in fact, occur. This possibility was suggested by some informally noted aspects of *Ss'* behavior in the Tarpy and Glucksberg study,[7] where the IT consisted of reading digits to the beat of a metronome. On a number of occasions, a consonant from the CCC trigram was substituted for an IT digit. *Ss* often fell behind the metronome beat, and IT errors did occur. Unfortunately, this earlier study was not designed to permit reliable assessment of IT performance, and so incentive effects as a function of IT performance could not be examined.

The present study was designed to examine incentive effects upon short-term retention as a function of IT performance. When IT performance is perfect, it is assumed that no covert rehearsal occurs. When IT performance is inaccurate, *i.e.* when *Ss* are inattentive, it is assumed that opportunities for covert rehearsal exist. An appropriate IT for this purpose is one in which performance is essentially error-free when *Ss* do not engage in covert rehearsal, *i.e.* when they pay full attention to it. If incentive effects are obtained irrespective of IT performance, *i.e.* irrespective of opportunity for covert rehearsal, then Weiner and Walker's arousal hypothesis is supported.[8] If incentive effects are found only when IT errors occur, *i.e.* only with opportunity for

[5] B. Weiner, Motivation factors in short-term retention: II, Rehearsal or arousal? *Psychol. Rep.*, 20, 1967, 1203–1208.
[6] Weiner, *op. cit.*, *Psychol. Rep.*, 20, 1967, 1203–1208.
[7] Tarpy and Glucksberg, *loc. cit.*
[8] Weiner and Walker, *loc. cit.*

covert rehearsal, then a differential-rehearsal interpretation of motivational effects upon short-term retention is supported.

Method

Subjects. Ss were 97 experimentally naive undergraduates (men) attending Princeton University. These Ss served as paid volunteers.

Materials and Procedure. (*1*) *Interpolated task.* Rehearsal-preventing activity consisted of keeping track of the number of times the first digit of a 20-digit series occurred in that series. The digits were tape-recorded at the rate of 2 per sec. to fill a 10-sec. retention interval. Task difficulty was defined in terms of the mean percentage of first-digit repetitions. An easy IT was used with additional, more difficult IT series included as 'dummy' trials to insure that the *Ss* were alert and could not be correct by guessing. The various 20-digit series were randomly determined except that half the series began with even digits, half with odd. *(2) Recall task.* The stimuli were 24 CCC trigrams with a mean association value of 30%.[9] The recall stimuli, like the IT digits, were tape-recorded at a rate of 2 per sec. The sequence of events constituting a trial was: a 1-sec. buzzer ready-signal, followed 5 sec. later by the CCC trigram; this was followed immediately by the 20 IT digits. At the end of the series of digits, *Ss* had 10 sec. to write the trigram and the count of first-digit occurrences. For example, if a trial sequence consisted of: *SBM13192417311262847221*, then the correct answer would be *SBM 6*, where the number *6* refers to the number of times the first digit, *1*, appeared in the 20-digit series. Two incentive conditions, high (H) and low (L), were employed. Thirty-six *Ss* were told that if both trigram and digit-count were correct, then they would be paid 10¢ when the first digit was odd, and 1¢ when the first digit was even. For 39 *Ss*, the incentive cues were reversed, *i.e.* H when the first digit was even, L when it was odd. There were four H and four L trials plus 16 'dummy' trials (SH and SL) with greater IT difficulty. *Ss* were tested in groups of 10 to 20. Trigram-recall performance was scored according to Weiner's method[10] of giving one point for each correct consonant and one point for each correct position, yielding a maximum score of six points for any given trial. *(3) Control conditions.* Two groups of 11 *Ss* each served as controls for (*a*) trigram retention when the IT stimuli were presented but were to be ignored and (*b*) IT performance when the trigrams were presented but were to be ignored. Identical stimuli were presented to all groups.

Results and Discussion

Control Groups

Trigram retention. Performance was virtually perfect when *Ss* were instructed to recall the trigram and to pay no attention to the IT. Only six errors were made in 264 trials (11 *Ss* given

[9] L. R. Witmer, The association value of 3-place consonant syllables, *J. genet. Psychol.*, 47, 1935, 337–360.
[10] Weiner, *op. cit., Psychol. Rep.*, 20, 1967, 1203–1208.

24 trials each). These data demonstrate that *Ss* could accurately hear the tape-recorded stimuli and could recall them with ease over a 10-sec. interval in the absence of interpolated activity.

Digit count. When *Ss* were given both the recall stimulus and the IT but were instructed to ignore the stimulus, errors on the IT were appropriately low. Numerous errors, however, did occur on the more difficult IT 'dummy' trials, indicating that any incentive effects (on trigram retention) would be uninterpretable for those trials.

Experimental Groups

Recall performance regardless of IT performance. Mean recall scores (expressed as a percentage of the total possible score) are presented in Fig. 1 for both incentive conditions. A Wilcoxon matched-pairs, signed-ranks test indicated no significant difference between the groups.[11] The failure to obtain an incentive effect is contrary to previous findings.[12] There are several possible reasons for this failure to replicate the incentive effect. The IT task used in the present study was different from Weiner's,[13] and, unlike Weiner's procedure, *Ss* here had to perform accurately on both the recall task and the IT; they could not sacrifice IT performance for recall performance. According to the covert-rehearsal hypothesis, incentive effects should be obtained only when IT performance is so sacrificed, *i.e.* when covert rehearsal occurs. For data bearing directly upon this issue, recall performance as a function of incentive and IT performance must be considered.

Recall performance as a function of IT accuracy. In order to evaluate recall performance as a function of incentive when IT performance was accurate (IT+), only those *Ss* who had both high and low incentive trials under IT+ conditions were included in the analysis (Fig. 1). Similarly, only those *Ss* who displayed inaccurate (IT−) performance on both high and low incentive trials were chosen (Fig. 1). Each *S* was assigned a mean recall score for the relevant entry so that each *S* would contribute equally to the group means. A Wilcoxon matched-pairs, signed-ranks test

[11] Nonparametric statistics were used since the maximum score of six points would tend to limit the upper portion of a distribution.

[12] M. I. Posner, Components of skilled performance, *Science*, 162, 1966, 1712–1718; Weiner, *op. cit., Psychol. Monogr.*, 80, 1966, No. 18, Whole No. 626.

[13] Weiner, *op. cit.*, *Psychol. Rep.*, 20, 1967, 1203–1208.

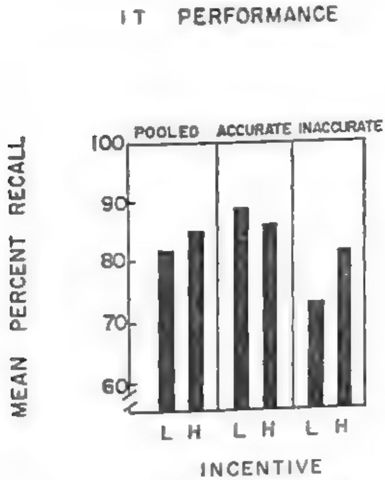


FIG. 1. MEAN RECALL SCORES AS A FUNCTION OF INCENTIVE AND IT PERFORMANCE

was then used to evaluate the effect of incentive. In the IT+ condition ($N = 66$) where, presumably, opportunities for covert rehearsal were greatly reduced, performance was not affected by incentive. In fact, slightly better performance was observed under low incentive than under high, although this difference was not significant. In the IT− condition ($N = 46$), where, presumably, covert rehearsal may have occurred (*S*s were explicitly inattentive), recall performance was significantly superior under high incentive, $z = 2.00$, $p < .05$. In summary, the results indicate that recall was facilitated by high incentive only when *S*s had the opportunity to rehearse. The findings reported by Weiner[14] are presumably due to the lack of control (or measurement) of such rehearsal opportunities.

In addition to the failure of the present data to lend empirical support to the arousal hypothesis, the hypothesis can be questioned on theoretical grounds. It is difficult to conceive of an arousal mechanism directly influencing trace processes in a discrete trial-to-trial fashion in the within-*S*s designs commonly employed.[15] The intertrial intervals are typically short, on the order of 10–30 sec. A state of high arousal would be expected to carry over from one trial to another, and this should mask any arousal effects that might be operative. On both empirical and logical

[14] Weiner, *op. cit.*, *Psychol. Monogr.*, 80, 1966, No. 18, Whole No. 626.
[15] Weiner, *op. cit.*, *Psychol. Monogr.*, 80, 1966, No. 18, Whole No. 626.

grounds, then, a covert-rehearsal interpretation of incentive effects is strongly suggested.

One final implication of the data obtained in this study merits brief comment. In general, data from short-term memory experiments employing variants of the Peterson and Peterson technique[16] have been obtained under procedures which do not provide for assessment of interpolated activity performance. To the extent that interpretation of these data rests on the assumption that covert rehearsal has been prevented, such interpretations may be in error. It would seem important, in light of present data, to evaluate the extent to which a rehearsal-preventing task does, in fact, prevent rehearsal. An interpolated task of the type employed here would seem to provide a technique for assessment of covert rehearsal and may profitably be used where prevention of covert rehearsal is an integral part of an experimental design.

Summary

Incentive effects upon short-term retention of CCC trigrams were examined in an experimental paradigm which permitted evaluation of the extent to which *Ss* paid attention to a rehearsal-preventing interpolated task. When interpolated task performance was accurate, it was assumed that covert rehearsal did not take place. When interpolated task performance was inaccurate, opportunity for covert rehearsal was presumed to be present. Incentive facilitated short-term retention only in this latter condition, supporting a covert-rehearsal interpretation of previously reported incentive effects upon short-term retention. General methodological implications for the study of short-term retention are briefly noted.

[16] Peterson and Peterson, *loc. cit.*

NOTES AND DISCUSSION

A METHODOLOGICAL STUDY OF INCIDENTAL AND INTENTIONAL LEARNING

When Saltzman studied the role of the orienting task in incidental and intentional learning, his procedure differed from most in that he used an orienting task (card-sorting) that consisted of three incidental presentations of the stimulus material.[1] After all his *Ss* had been thus exposed to the stimulus items, half were instructed to learn the material (intentional learners) while the other half were asked merely to observe the stimulus items (incidental learners) during the three card-sorting trials that then followed. Saltzman found that with equal initial exposure to the material, the groups learned at the same rate. He concluded that the greater efficiency usually reported for intentional learning resulted from the intentional *Ss'* greater initial experience with the stimulus material. Experiment I in the present study was an attempt to replicate and extend Saltzman's findings and to control for three factors which, in Saltzman's study, allowed other interpretations of his results. The uncontrolled agents of concern were these: *(1) Postorientation testing.* Saltzman tested his intentional *Ss*, but not his incidental *Ss*, following the orienting task. Their generally low performance on this test could have lowered their motivation, thus eliminating the differential effects of instructions to learn.[2] *(2) Sorting rate.* Saltzman's intentional learners worked for significantly longer periods than did his incidental learners. This prolonged practice could have generated fatigue or boredom which depressed the intentional learners' scores. *(3) Exposure time.* The face-upward card-sorting that Saltzman's *Ss* did left

* Received for publication September, 1968. The research reported was supported by the Air Force Technical Training Branch, Behavioral Sciences Laboratory, under Project 7183, Psychological Research on Human Performance. Dr. Neiberg is now at the University of Michigan, Flint; Dr. Levine is now at the State University of New York at Stony Brook.

[1] I. J. Saltzman, The orienting task in incidental and intentional learning, this Journal, 66, 1953, 274–277.

[2] For evidence that motivation is an important contributor to both incidental and intentional learning, see Ronald Johnson and Calvin Thomson, Incidental and intentional learning under three conditions of motivation, this Journal, 74, 1962, 284–288.

some of the stimulus items exposed for long periods of time. Incidental *S*s could benefit from such exposure, but intentional *S*s could not, since before a card was sorted, the intentional *S* had the opportunity to study the card for as long as he deemed necessary. This additional benefit for the incidental *S*s could have diminished any differences which existed between the groups. Experiment I of the present study thus attempted to replicate Saltzman's work while assessing the contribution of post orientation testing, sorting rate, and exposure time to such a comparison of incidental and intentional learning. Experiment II was then concerned with the retention of the incidentally learned material. A long-held hypothesis is that incidental learning may occur under some circumstances but that material so acquired is not retained as well as intentionally learned material. Experiment II tested this hypothesis by comparing retention of incidentally and intentionally learned material after a 24-hr. rest.

EXPERIMENT I

Subjects. The *S*s were 130 college students from an introductory psychology course.

Apparatus and materials. Three identical sets of 32 cards were prepared.[3] Each card was 2½ × 2¼ in. with a two-digit number (2 × 2 in.) lettered on the face. A Johnson card-shuffler (Model 50) was used to insure that at least one well-shuffled deck of cards was available at all times. Of the 32 cards, eight contained odd numbers greater than 50, eight contained even numbers greater than 50, eight contained odd numbers less than 50, and eight contained even numbers less than 50.

Design. Six groups were employed, with these conditions: Group I, same as Saltzman's intentional learners; Group II, same as Saltzman's incidental learners; Group III, same as Group I but without postorientation testing; Group IV, same as Group II but with instructions to sort for accuracy rather than speed; Group V, same as Group III but with instructions to look away after sorting (close proctoring assured that these *S*s complied with these instructions); and Group VI, same as Group IV but with instructions to look away after sorting. The objectives of the experiment were achieved by use of these groups in the following ways: (1) A test of the replicability of Saltzman's findings was possible by comparison of Groups I and II. (2) An evaluation of the affect of postorientation testing on intentional learning scores was obtained by comparison of Groups I and III. (3) The effects of sorting rate on incidental and intentional learning scores were determined by comparison of Groups III and IV; this was possible since instructions to sort for accuracy rather than speed resulted in a slower sorting rate for incidental learners.

[3] These cards originally were blank playing cards contributed to the research effort with the compliments of the U. S. Playing Card Co., Cincinnati, Ohio.

(4) The effects of exposed stimuli, instructions to learn, and an interaction of these were studied by comparison of Groups III, IV, V, and VI.

Procedure. First, all *Ss* were given three orienting trials with instructions to sort the 32 cards, as rapidly as possible, into the four groups of odd or even, above or below 50, numbers. Next, *Ss* in Group I were given a 75-sec. test to see how many numbers they had learned; during this time, *Ss* in the other groups rested. Then, all *Ss* were instructed to continue sorting for three more trials. Those *Ss* assigned to the incidental groups (II, IV, and VI) were told to sort just as they had earlier. In addition, *Ss* in Groups IV and VI were instructed to sort for accuracy rather than speed, and *Ss* in Group VI were instructed to look away after sorting. Those *Ss* assigned to the intentional groups (I, III, and V) were asked to try to learn the numbers while they were sorting in this part of the experiment. Group V *Ss* were further instructed to look away after sorting. Finally, all *Ss* were given a 75-sec. recognition test. A mimeographed sheet with the numbers 11–99 was given to each *S* with instructions to circle as many of the numbers which had appeared on the cards as he could within the allotted time.

Results and discussion. The number of correct responses minus 80% of the number of incorrect responses served as a corrected performance score. The mean performance and variance for each group appear in Table I. A t test was performed between the two groups modeled after Saltzman's (Groups I and II). The differences found were not statistically significant (t = 1.64; df = 38), indicating the success of the present experiment in replicating Saltzman's work. These nonsignificant differences imply that formal instructions do not lead to more efficient learning as long as equal exposure to an orienting task is provided.

Saltzman's findings thus duplicated, the possibility that post-orientation testing depressed intentional learning was examined by comparing the intentional-learning group which had been so tested (Group I) with the intentional-learning group which had not been so tested (Group III). No significant differences were found (t = 0.13; df = 38), indicating that the testing experience probably had no harmful effect on criterion test scores. As a further test, Groups II and III were compared. These were composed, respectively, of incidental and intentional learners, none of whom were tested after the orienting task. The lack of a sig-

TABLE I

MEAN RECOGNITION SCORES AND VARIANCES: EXPERIMENT I

	Group					
	I	II	III	IV	V	VI
Mean	12.0	9.5	12.2	9.4	12.6	10.2
Variance	15.0	29.1	34.2	22.2	20.4	11.5

STIMULUS VARIATION AND FREE RECALL: A CONFIRMATION

Several recent experiments have explored the roles of repetition and variety in learning and recall.[1] Bevan, Dukes, and Avant, for example, reported three experiments in which stimulus variety was the major independent variable.[2] In their Experiment I, the same *Ss*, when presented with slides depicting common objects, recalled more objects when the generic class was represented by two different specimens (AB) than when the class was represented by one specimen presented twice (AA). In Experiment II the results were repeated using nouns for which the adjectival modifier was either changed or duplicated. Experiment III used five degrees of variation of pictorial material, with each group of *Ss* receiving a different degree of variation. Again, the superiority of variety was demonstrated, but with some evidence of a saturation effect. That is, classes represented by four different specimens (ABCD) were no better recalled than classes represented by two identical and two different specimens (AABC). These authors suggested that the results may be due either to enhanced attention to unduplicated stimuli or to enhanced probability of recall due to increased associations to the class when the specimens were varied.

Perhaps one way to rule out one of the two hypotheses would be in terms of the between-groups vs. within-*Ss* designs. If enhanced attention were the major factor, then one might expect the enhancement from varied stimuli to also exert an effect on the duplicated stimuli in a design in which each *S* received experience under both duplicated and varied conditions. The effect of variety would thus be minimal compared to the situation where each *S* experienced only one level of variety. Since the Bevan, Dukes, and Avant report used both a within-*Ss* design (Experiment I) and a between-groups design (Experiment III), and obtained the effect, it would appear that the general-attentiveness hypothesis is weakened. The present experiment aimed at confirming this effect in a within-*Ss* design using three levels of

* Received for publication August 15, 1968.

[1] William Bevan, William F. Dukes, and Lloyd L. Avant, The effect of variations in specific stimuli on memory for their superordinates, this JOURNAL, 79, 1966, 250–257; William F. Dukes and William Bevan, Stimulus variation and repetition in the acquisition of naming responses, *J. exp. Psychol.*, 74, 1967, 171–181; William Bevan and William F. Dukes, Stimulus-variation and recall: The role of belongingness, this JOURNAL, 80, 1967, 309–312.

[2] Bevan, Dukes, and Avant, *loc. cit.*

variety. The use of three, rather than two, levels of variety was also intended to provide information on the location of a saturation point for variety in the within-*S*s design.

Method

Stimulus materials. Stimulus materials consisted of 72 color slides of common objects. The slides were divided into three groups as follows: (1) AAAA—Six classes of objects, each represented by four different specimens. Each specimen was presented four times. The objects in this group were guns, toys, candles, books, scissors, and maps. (2) AABB—Six classes of objects, each represented by two different specimens, with each specimen presented twice. Classes were bridges, houses, drums, beds, women, and cars. (3) ABCD—Six classes of objects, each represented by four different specimens, with each specimen presented once. Classes were children, desserts, clocks, cameras, flowers, and cats.

Subjects. *S*s were 42 students (20 men, 22 women) in two introductory psychology classes at Georgia State College. They were tested in the classroom during the first few minutes of the class hour.

Procedure. The 72 slides were loaded in random order into the magazine of a Kodak Carousel projector. *S*s were each given a sheet of paper and then instructed to look at the slides without saying or writing anything. The slides were presented for 2 sec. each. The interstimulus interval was the time required for the Carousel to cycle. After all slides were shown, *S*s were allowed 3.5 min. to recall, as instructed, as many of the objects as they could.

Results

The number of objects recalled was determined and was translated into the number of categories correctly recalled. In some instances more than one object in a given category was recalled, and in such cases only one correct response was counted. In general, any response indicating correct recall of the category was accepted. For example, for the category 'bed,' acceptable responses were 'bedroom suite,' 'bed,' 'bedroom,' or anything indicating that the *S* recalled a situation in which a bed was present.

The mean numbers of categories recalled (out of a possible total of six) for each of the three conditions were: AAAA, 3.2; AABB, 4.1; ABCD, 4.5. The difference between means was significant by a Friedman two-way analysis of variance ($\chi^2_r = 23.3, p < .001$). One-tailed Wilcoxon matched-pairs, signed-ranks tests on adjacent means indicated that each mean differed from its neighbor: AAAA vs. AABB, $N = 26$, $z = 3.50$, $p = .0002$; AABB vs. ABCD, $N = 28$, $z = 1.85$, $p = .0322$. Thus, variety exerted a significant effect upon recall. In order to test for saturation effect, the relative size of the

differences between adjacent means was assessed with the Wilcoxon test. Although the AAAA vs. AABB effect appeared greater than the AABB vs. ABCD effect, the difference was not significant ($N = 33$, $z = 1.51$, $p > .05$). Thus, a reliable saturation effect was not evident in these data.

Discussion

The present results offer confirmation of the Bevan, Dukes, and Avant findings previously reported, in that variety was shown to enhance recall. In addition, since the effect was obtained between three increasing levels of variety in a within-*S*s design, they lend further support to the notion that the effect is not due solely to a general enhancement of attentiveness in the conditions with greater variety. Whether or not the effect is due to the increased number of associations available to the more varied stimuli remains to be seen. One could equally well argue that responsiveness to the duplicated stimuli is somehow inhibited due to the operation of a pooling process such as is hypothesized by adaption-level theory.[3] Further research in which the relative roles of variation in contextual and focal stimuli are assessed, and in which the association value of stimuli presented is manipulated, should help to explain the phenomenon. Finally, it should be noted that since the present study did not indicate a saturation point for variety, further research should be performed to determine the limits of this effect, especially in the within-*S*s design.

Georgia State College

WALTER F. DAVES
MARY ADKINS

[3] Harry Helson, *Adaptation Level Theory: An Experimental and Systematic Approach to Behavior*, 1964.

A REPORT ON MYERS' NEW AUTOKINETIC PHENOMENON

In 1959 Myers reported a new autokinetic phenomenon.[1] Nine out of 10 of his *S*s reported that a thin luminous line appeared to break up into dots and dashes when viewed in a dark

[1] R. D. Myers, A new autokinetic illusion, this JOURNAL, 72, 1959, 140–141.

room. In addition, six *Ss* observed "irregular movement, drift, or flow" among the dots and dashes.[2] The target was a 35-in. wooden dowel (0.5 in. in diameter), painted longitudinally on one half of the surface with green luminous paint. This dowel could be rotated on its longitudinal axis, presenting less and less of the luminous paint, until just a thin, horizontal line was visible to *S*. Due to the unique aspects of this target (narrow width and luminous source), there was some question in the mind of the present investigator as to the reliability of the phenomenon. Luminous paint that is charged by a light-source shows some variability in rate of decay. This sometimes results in scintillation of the luminous object. In view of this factor, the present study replicated the phenomenon with a more conventional stimulus source and tested the effect of target width.

Method

Procedure. A preliminary study indicated that the phenomenon did occur when a luminous target similar to Myers' was used. A more conventional target was then constructed, which consisted of an aperture covered with translucent paper (draftsman's tracing paper). Behind and below this aperture was a tubular filament light (Westinghouse Lumline, 60 w.) of adjustable brightness. The width and length of the aperture could be adjusted between trials by *E*. The effect of target width on *S*'s report of the phenomenon was tested in this manner: The target was presented horizontally at eye level, 10 ft. from *S*, and was kept at a constant brightness. Target widths of 1/16 in., 1/8 in., 1/4 in., 3/8 in., 1/2 in., and 5/8 in. were presented to each *S* in either ascending (increasing width) or descending (decreasing width) series using the method of limits. *Ss* were instructed to fixate on the center of the target and to report what they saw.

Results

Of the 44 *Ss*, 14 *Ss* (32%) reported segmentation of the target. Typical descriptions included: "It looks like Morse code" and "It seems to be separated into small segments." Again of the 44 *Ss*, 30 *Ss* (68%) reported dots or "specks" flowing within the line. Nine of the *Ss* (20%) reported both segmentation and dots. The reports of dots or segmentation fell off rapidly with increased target width. No reports of either were made at widths beyond 1/2 in., and only two reports were made at 1/2 in. Contrary to Myers' report, 30 *Ss* (73%) also reported normal autokinesis or right-left, up-down, and bending movement of the entire line;

[2] Myers, *op. cit.*, 140

Myers indicated that none of his *Ss* reported normal autokinesis if the new phenomenon was reported.

Summary

It thus appears that the phenomenon reported by Myers can be replicated and is not merely a function of the luminous target. Not only is the phenomenon reported with a target produced by incandescent illumination, but target width affects the phenomenon. A review of autokinetic literature indicates no adequate explanation. The practical significance of the phenomenon is questionable, however, since it evidently had gone unnoticed until Myers' report.

Berea College J. F. White

C. LLOYD MORGAN AND THE THEORY OF INSTRUMENTAL LEARNING

About 75 years ago, the English psychologist C. Lloyd Morgan began to study what he then called "trial and error" learning and what we today call "instrumental" learning.[1] His subjects were domestic animals—chicks and dogs. Morgan was an evolutionist and certainly not inclined, as he put it, to set a gulf between the minds of animals and men, but he was tired of the exaggerations of the anecdotalists, who, in their determined efforts to prove Darwin's theory, seemed to have lost all sense of proportion. It was clear to Morgan that the time for anecdotes had passed and that the time for experimental analysis had come. To explain the changes in behavior which he observed, Morgan assumed that his animals learned, by a process of association, to anticipate the consequences of their actions and that they then behaved in accordance with the pleasantness or unpleasantness of those anticipated consequences. Morgan's conception of instrumental learning is strikingly similar to our own and worth examining in some detail.

* This paper was prepared for a symposium on the history of comparative psychology at the 1967 meetings of the American Psychological Association.

[1] C. L. Morgan, *Introduction to Comparative Psychology*, 1894, 1–382.

Reproduced in Fig. 1 is a diagram drawn by Morgan in 1894 to show how the general tendency of newly-hatched chicks to peck at small objects is modified by experience.[2] To begin with, Morgan observed, the chick pecks indiscriminately at a variety of objects, some edible (which it swallows), others inedible (which it rejects). As time goes on, the chick pecks at the edible objects more and more readily, pecks at the inedible objects less and less readily, and in the end avoids the inedible objects altogether. The *1* in the diagram represents a visual stimulus from a given object which reflexively elicits pecking via a built-in, lower-lying pathway. *M.R.* represents the pecking response. The visual stimulus also activates a higher visual center, which is labeled *V*. The *2* in the diagram stands for kinesthetic feedback from the pecking response, which activates a cerebral center for kinesthesis, labeled *M*; and *3* stands for a gustatory stimulus from the pecked object, which activates a cerebral center for taste, labeled *T*. The contiguous activation of these three cerebral centers produces new connections among them, which are indicated in the diagram by appropriate lines. As the new connections are strengthened by repetition, the visual stimulus alone comes to activate the kinesthetic center and the taste center, *i.e.* upon seeing the object, the animal remembers the movement and the taste.

Now there is the problem of how these memories—or the connections that mediate them—affect the behavior of the animal, because it is change in behavior which the theory purports to explain. Guthrie later could say of Tolman that he left the animal "buried in thought,"[3] but the same criticism cannot be made of Morgan. In the early stages of instrumental learning, Morgan assumed, behavior is regulated by cerebral control centers. The control center for pecking, labeled *C.C.* in the diagram, is aroused to begin with by the kinesthetic center for pecking via a built-in connection. The control center is connected with the lower-lying motor centers for pecking by a descending pathway, labeled *P.T.* (which stands for *pyramidal tract*). The function of the control center is to facilitate or to inhibit pecking, and which of these opposed effects it has will depend upon the prevailing hedonic tone. If the remembered taste is pleasant, pecking will be facilitated; if it is unpleasant, pecking will be inhibited. In the early stages of learning, then, the control center

[2] Morgan, *op. cit.*, 183.
[3] E. R. Guthrie, *The Psychology of Learning*, 1935, 172.

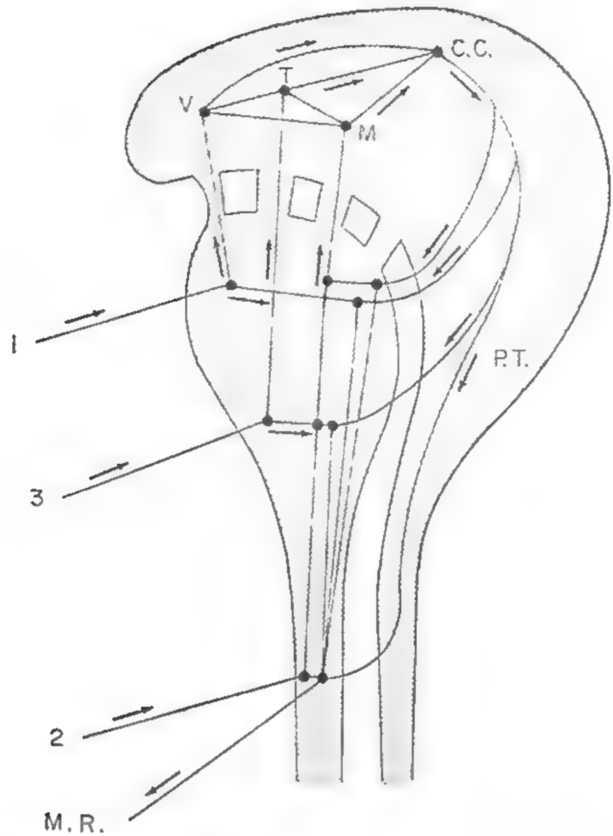


FIG. 1. MORGAN'S EXPLANATION OF THE DEVELOPMENT OF CEREBRAL CONTROL OVER THE PECKING RESPONSE OF THE CHICK

(*1*, visual stimulus; *2*, kinesthetic feedback from pecking; *3*, gustatory stimulus; *M.R.*, pecking response; *V*, *T*, and *M*, cerebral centers for vision, taste, and kinesthesis; *C.C.*, cerebral control center for the pecking response; *P.T.*, pyramidal tract.)

plays a rather important role, but after a time its role may become much less important. Consider what happens in the course of extensive experience with a pleasant-tasting object. Since the control center facilitates pecking, the lower-lying visual and motor centers are contiguously activated with each presentation of the object. Their contiguous activation strengthens the connection between them, which becomes strong enough eventually to bypass the cerebral control mechanism altogether.

The pecking example is a rather special one, of course, because it involves a preexisting response tendency which is simply strengthened or weakened by experience, but the theory is perfectly general. It will also explain how a dog learns to escape

from an enclosure by lifting a latch, another instance of instrumental learning studied by Morgan. In the latch example, no preexisting connection between the sight of the latch and the lifting response need be assumed. According to Morgan's theory, whenever—and for whatever reason—an animal makes a response in the presence of some stimulus, S_1, and that response produces another stimulus, S_2, the cerebral centers for S_1, S_2, and the kinesthetic feedback from the response are contiguously activated. As a result of their contiguous activity, the three centers become interconnected, and it therefore becomes possible for S_1 alone to activate the center for S_2 and the control center for the response. The control center then activates or inhibits the response, depending upon the incentive value of S_2. When S_2 is pleasant, the response is made each time it is presented, and a new connection gradually develops between the lower-lying sensory and motor centers, even if there was none to begin with. The new S-R connection gains in strength, and eventually it becomes capable of mediating a highly automatic kind of behavior in which the higher processes of anticipation and evaluation of the consequences of response play no part.

Is it surprising to find such a theory in the literature of 1894? In some respects, certainly, the theory is primitive, but there are theories no less primitive coming today out of eastern Europe, where instrumental learning is being rediscovered. Our own thinking is much more sophisticated, of course, and it has a vastly broader empirical base, but our fundamental ideas about instrumental learning are much the same as Morgan's. Like Morgan, we distinguish between learning and performance. Like Morgan, we accept contiguity as a sufficient condition for learning. Like Morgan, we assume that animals learn to anticipate the consequences of their actions and that their anticipations guide their actions. All these ideas seem quite reasonable to us today. As for Morgan, trained in associationism and in biology, setting out to deal with the behavior of his animals in terms of functional categories derived from human experience, but intent upon the most parsimonious employment of those categories, it is difficult to imagine how he could have arrived at any other interpretation. His learning principle was that of his contemporaries, and the assumption on which he depended for the translation of learning into performance—the assumption of preexisting connections between the proprioceptive centers for skeletal responses and the

motor centers for those responses—was in the air of the time. One finds it, for example, in James.[4]

In my opinion, Thorndike's interpretation of instrumental learning was much the more inventive.[5] Thorndike took his experimental procedures directly from Morgan, and his basic findings were the same, but his interpretation of those findings was boldly original. It is not difficult to understand that Thorndike's S-R reinforcement principle should have aroused so much interest or that so much effort should have been devoted to an examination of its possibilities and implications. What is surprising is only that it should have taken us so long to become aware of the limitations of the principle and that so much hostility to the earlier view should have been engendered in the process. The learning-performance distinction was not generally accepted for 40 years. It was Tolman who showed the need for the distinction,[6] and it was Hull whose acceptance of the distinction gave it scientific respectability,[7] although neither Tolman nor Hull was much concerned with the mechanisms by which learning is translated into action. Hull continued, of course, to cling to the reinforcement principle, whose main formal advantage once had seemed to lie in the fact that it required no supplementary activating mechanism. In the next 15 or 20 years, the reinforcement principle was eroded slowly but inexorably, and expectancy (in a variety of linguistic guises) came gradually to the fore.

The extent of the change in our thinking between Hull's day and the present may be difficult to appreciate for two reasons. One is that the change has been so gradual. The second is that, because formal learning theory has fallen into disrepute, we have no really adequate statement of the assumptions prevailing today. If we had such a statement, its difference from Hull's statement of 1941, and its similarity to Morgan's statement of 1894, would be at once apparent. With some anatomical updating, and a few changes in vocabulary, Morgan's statement would, in fact, well represent the modal view of the present.[8] The history of our work

[4] See G. A. Kimble (*Foundations of Conditioning and Learning*, 1967, 147–153) for an analysis of James's early attempt at a learning model.

[5] E. L. Thorndike, Animal intelligence: An experimental study of the associative processes in animals. *Psychol. Rev. Monogr. Suppl.* 2, 1898, 1–109.

[6] E. C. Tolman, *Purposive Behavior in Animals and Man*, 1932, 1–463.

[7] C. L. Hull, *Principles of Behavior*, 1943, 1–422.

[8] See, for example, N. E. Miller, in M. R. Jones (ed.), *Nebraska Symposium on Motivation*, 1963, 94–96; F. D. Sheffield, in W. F. Prokasky (ed.), *Classical Conditioning*, 1965, 316–321.

on the problem of instrumental learning in the years since Morgan might well be characterized as a systematic and determined but unsuccessful effort to find an acceptable alternative to Morgan's view. I do not mean to imply, of course, that we have done nothing in those years but return to our starting point. It is one thing to propose a feasible interpretation of a phenomenon in the first stages of its investigation, and quite another to return to that interpretation many experiments later, after earnest examination and rejection of a variety of alternatives. Nor do I mean to imply that we have nowhere else to go. Let me point out, briefly, one new direction.

Until the present time, the main subjects of our research on instrumental learning have been the rat and a few other mammals, which we have tended to treat as representative of animals in general. It is the data for these animals that seem now to require a contiguity rather than a reinforcement principle. In recent years, however, the intensive study of still simpler animals has begun to yield data for which Hull's theory provides a better fit than Morgan's. If I had to choose the one kind of experiment which more than any other undermined our confidence in the reinforcement principle, it would be the Crespi experiment.[9] Its procedure is simple. Two groups of rats are trained in spaced trials to make a simple instrumental response, one group with low reward and one group with high reward; then half the low-reward animals are shifted to high reward, and half the high-reward animals are shifted to low reward. The Hullian prediction is illustrated in Fig. 2. In the first part of the experiment, the response latencies of the two groups should fall to different asymptotic levels, lower for the high-reward (40-unit) animals than for the low-reward (4-unit) animals, because high reward produces greater habit strength than does low reward. With the shift from low to high reward (4–40 units), the performance of the low-reward animals should improve gradually, because their habit strength increases gradually. The shift from high to low reward (40–4 units) should have no effect on the performance of the high-reward animals, because their habit strength has already reached the high-reward asymptote.

Crespi's results for the rat were, of course, very different from these. Both groups showed sharp contrast effects—the so-

[9] L. P. Crespi, Quantitative variation of incentive and performance in the white rat, this JOURNAL, 55, 1942, 467–517.

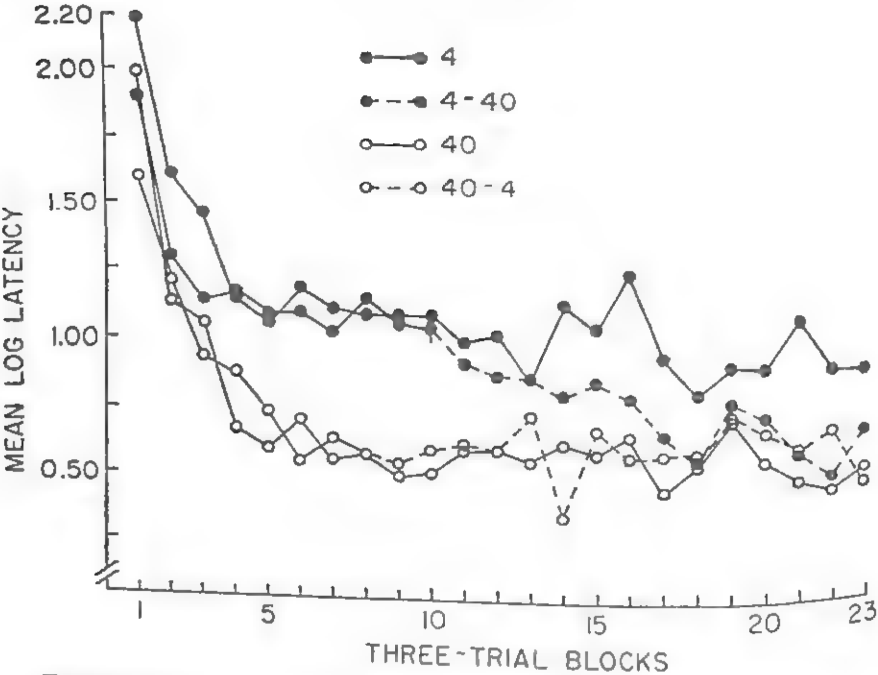


FIG. 2. LATENCY OF A SIMPLE INSTRUMENTAL RESPONSE AS A FUNCTION OF AMOUNT OF REWARD

(The two amounts of reward were 4 and 40 units. The curves labeled 4–40 and 40–4 show the effects on latency of change in amount.)

called elation and depression effects—which suggested that the animals were reacting emotionally to the changes in amount of reward. But how could the animals have reacted to the change in amount if they had no memory of the previous amount? On the basis of Crespi's results, Hull abandoned the assumption that habit strength increases with amount of reward.[10] High-reward animals respond more rapidly than low-reward animals, he decided, not because they have greater habit strength, but because they anticipate a higher reward. Hull did not give up the reinforcement principle altogether; in the final version of his system, reward still was necessary for learning, although the amount of learning was assumed to be independent of the amount of reward. That position was an awkward one, however, and a few years after Hull's death, Spence went over entirely to a contiguity interpretation of instrumental learning.[11]

The data of Fig. 2 come, not from an experiment with the rat,

[10] Hull, *A Behavior System*, 1952, 1–372.
[11] K. W. Spence, *Behavior Theory and Conditioning*, 1956, 1–262.

but from an experiment with the goldfish.[12] Unlike the data for the rat, they conform in every respect to the original Hullian prediction. What are we to make of them? One possibility is that the most fundamental processes of learning are different in fish and rat, that the reinforcement principle holds for the fish and the contiguity principle for the rat. This possibility seems to me rather unlikely. Another possibility is that *the reinforcement principle holds for both animals*—for the rat as well as for the fish—but that its operation in the rat is masked by expectational processes which have not yet appeared in the fish. This possibility is one which I am inclined to take quite seriously. Morgan's theory now is in the ascendance, but the last word of Hull and Thorndike may not yet have been heard.

Bryn Mawr College M. E. Bitterman

[12] Gilliam Lowes and M. E. Bitterman, Reward and learning in the goldfish, *Science*, 157, 1967, 455–457.

A TWO-CHAMBER HURDLE APPARATUS FOR USING ODORANTS IN AQUATIC ANIMAL-TRAINING

The apparatus described herein lends itself to various schedules of operant conditioning. While it was specifically designed to permit a more convenient way of using odorants as the S^D, it also permits better resolution of operant responses than the usual barrier-type hurdle box. The training tank described consists of a waterproof marine-plywood chamber 40 × 12 × 20 in. This chamber is divided into two compartments by a vertical divider 20 in. high. At the top of the divider is a plexiglass 'rocker platform' which is 18 × 12 × 4 in. and is driven by an electric motor (Superior SS50-P2-RC) mounted on the outside wall of the chamber. The motor is controlled by two sets of photoelectric

* Received for publication August 19, 1968. The design and construction of the electronic components of the apparatus reported were by Dr. Norman T. Welford, Director, Medical Electronics, University of Texas Medical Branch, Galveston, Texas. The project was supported in part by a grant from the Medical Research Foundation of Texas and in part by funds from institutional grants from the National Science Foundation, NSF-GU-484-CO-3, and the National Institutes of Health, 1SO1-FR-05427(06).

relay cells (Sigma Instruments 8RCO2A) mounted in the wall of the chamber, 4 in. on either side of the divider. Each photocell is connected to a cumulative recorder which permits automatic recording of each response on actuation of the photoelectric relay. The chambers contain copper wire-mesh grids (approximately 180 sq. in. each) on each side of the tank. Each side can be drained separately, manually or automatically with additional solenoids. The drain holes are 6 in. in diameter. Draining requires approximately 7 sec. when the tank is filled to the 18-in. level. Water in each compartment is continuously agitated by a 3½-in. rotor located on the floor of the chamber near the drain. The rotor is encased in a plastic cage and rotates at an average speed of approximately 200 rpm. Each chamber is equipped with an elevated drain pipe (stand pipe) which prohibits overflowing. Each chamber may also be fitted with a false plexiglass bottom in order to decrease the fishes' hurdle-response latencies.

As shown in Fig. 1, the rocker platform, constructed of crossbars, is mounted on top of the divider and has two alternate positions. When the fish swims into the rocker platform, it breaks the photoelectric beam which activates a relay and the platform tilts, lifting the fish out of the water and transferring it into the other compartment. The rocker platform remains in this altered position until the fish enters a second photoelectric beam which activates the drive motor, tilting the rocker platform and depositing the fish back in the first compartment. There is a time-delay mechanism in series with the photoelectric cells which gives the fish ample time to swim out of the rocker platform before he is flipped back into the other compartment.

The apparatus, as described, has several advantages over the usual barrier-type hurdle tank and simple target-type trainers. Because of its two-chamber design, odorants, which must be flushed out after each trial, may be used as CSs or S^Ds. Furthermore, because the fish must move a considerable distance, there is more topological definition in response measurements. In our laboratory, the apparatus has proven to yield excellent results in avoidance conditioning of fish 8–12 in. long. Small models could be built to accommodate smaller fish. And although the apparatus, as described, was designed for use of punishment and negative reinforcement, slight changes would make it satisfactory for use of positive reinforcers as well. Finally, we might note the recent findings by Garcia and his co-workers that gustatory

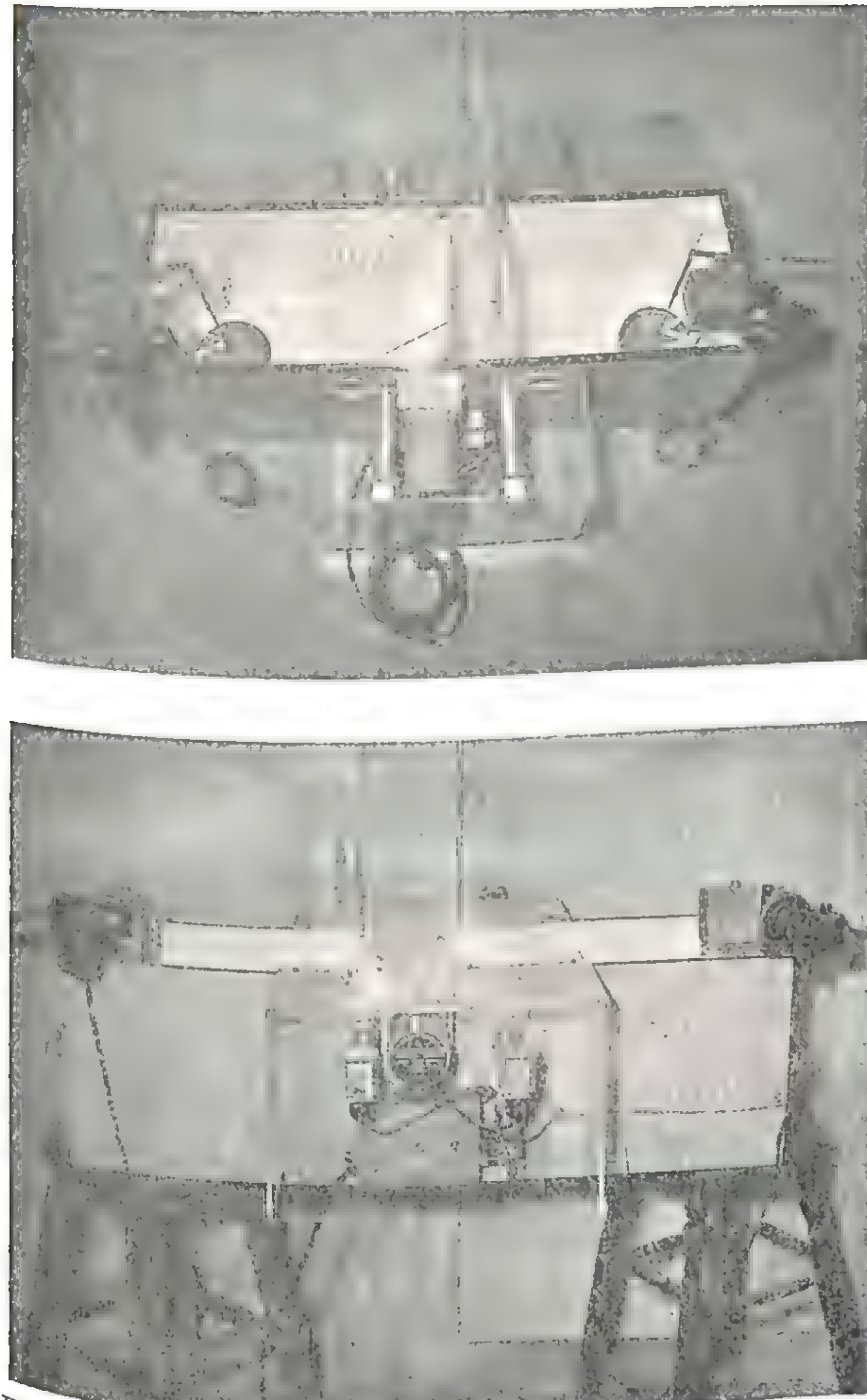


FIG. 1. A TWO-CHAMBER HURDLE APPARATUS FOR USING ODORANTS IN AQUATIC ANIMAL-TRAINING, TOP AND SIDE VIEWS

CSs do not become associated with noxious events in a manner typical of the classical paradigm[1]—findings which further indicate that taste stimuli readily become associated with 'internal' events and are not readily associated with 'external' events. Garcia and his co-workers, as well as most other investigators, have used Pavlovian paradigms. It would seem timely, and of great interest, to investigate the manner in which odorants operate in instrumental situations. The present paper provides an apparatus suitable for this purpose.

University of Texas Medical Branch,
Galveston, Texas

R. G. Benton
P. Maxcy, Jr.
B. Robertson
D. A. Rappoport

[1] J. Garcia, N. A. Buchwald, C. D. Hull, and R. A. Koelling, Adaptive responses to ionizing radiation. *Bol. Inst. Estud. med. Biol. Mex*, 22, 1964, 101–113; J. Garcia, N. A. Buchwald, B. H. Feder, R. A. Koelling, and L. F. Tedron, Ionizing radiation as a perceptual and aversive stimulus: I, Instrumental conditioning studies, in T. J. Haley and R. S. Snyder (eds.), *Response of the Nervous System to Ionizing Radiation*, 2nd International Symposium, 1964, 673–686.

A PROCEDURAL NOTE ON THE SIMULTANEOUS USE OF ELECTRICAL BRAIN STIMULATION AND GRID-DELIVERED SHOCK

In recent years, electrical stimulation of the brain (ESB) and grid-delivered electric shock have been used together in a number of experimental arrangements. Typical of these procedures has been the work of Brady and Conrad, Mogenson and Morrison, and Mogenson.[1] While the experimental objectives may vary, the standard grid-shock apparatus generates an electrical current

* Received for publication June 24, 1968. The research reported was performed under Contract DA-49-193-MD-2810 between the Office of the Surgeon General, United States Army, and the University of Maryland. The work was conducted according to the APA statement of "guiding principles for the humane care and use of animals," December 15, 1962.

[1] J. V. Brady and D. G. Conrad, Some effects of limbic system self-stimulation upon conditioned suppression, *J. comp. physiol. Psychol.*, 53, 1960, 128–137; G. J. Mogenson and M. J. Morrison, Avoidance responses to "reward" stimulation of the brain, *J. comp. physiol. Psychol.*, 55, 1962, 691–694; G. J. Mogenson, Avoidance responses to rewarding brain stimulation: Replication and extension, *J. comp. physiol. Psychol.*, 58, 1964, 465–467.

and unsystematically distributes it to pairs of floor grids in the experimental chamber. An often overlooked consequence of this arrangement is that even when shock is not being delivered, some of the grid elements are grounded. Therefore, when the animal is situated on the grid while receiving ESB, the amount of current being delivered may be somewhat altered since the animal is simultaneously grounded.

The solution to this problem involves the use of a double pole–double throw (D.P.D.T.) relay to disconnect *both* lines going from the shock generator to the scrambler (see Fig. 1) whenever the subject is receiving ESB. It should be noted that it is not sufficient to interrupt the 110-v. A.C. line to the shock generator. This may not disconnect the ground to the shock grids and may also introduce a great deal of inaccuracy into the internal timing mechanism which controls duration of the shock.

A second problem exists if the cranial electrodes are attached to the brain-stimulation unit at the time of grid-shock delivery. The presence of current at the electrode tip during grid-shock delivery has been repeatedly monitored on an oscilloscope in our laboratory. It was further noted in our investigations that the occurrence of current at this cranial locus was largely dependent on whether the electrodes were in contact with the stimulation unit at the time of grid-shock delivery. The presence of current at the electrode tip may seriously alter the behavioral properties of the grid shock. Brady and Conrad's failure to demonstrate conditioned suppression of responding reinforced by ESB may

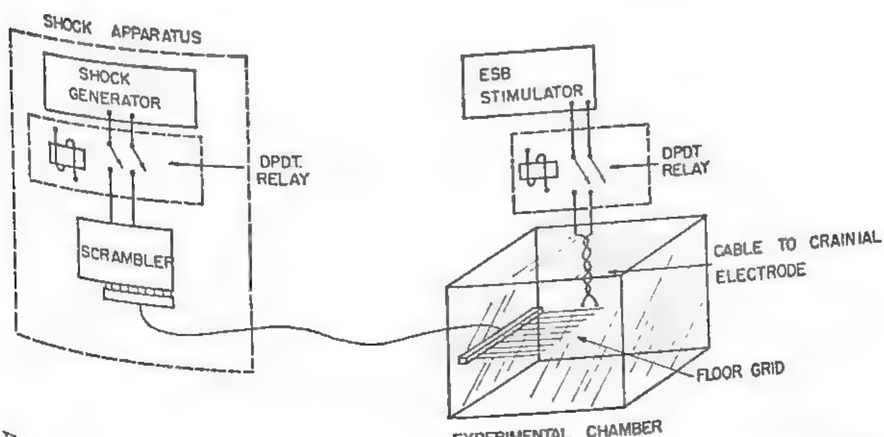


FIG. 1. CORRECT PLACEMENT OF D.P.D.T. RELAYS TO ISOLATE SHOCK APPARATUS FROM FLOOR GRIDS AND BRAIN-STIMULATION UNIT FROM CRANIAL ELECTRODES

be related to effects induced at the electrode tip by the grid shock.[2] Brady and Conrad also demonstrated that ESB-maintained response rate was accelerated following the delivery of grid shock, whereas food-reinforced responding was not.[3] The accelerative effects of grid shock in the ESB case may also have been related to grid-shock effects at the electrode tip.

A solution to this problem consists of electrically disconnecting the electrodes from the brain-stimulation unit by an additional set of D.P.D.T. relay contacts. Because of the capacitor-like properties of long cables, the relay which breaks the connection should be located as close to the subject as possible (see Fig. 1). A failure to locate the relay at this position may result in the same problem as if the relay were absent altogether.

California State College, Los Angeles
University of Maryland

HANK DAVIS
ROGER W. MCINTIRE
MELVIN KREITHEN
T. A. MCCULLOUGH

[2] Brady and Conrad, *loc. cit.*
[3] Brady and Conrad, *loc. cit.*

BOOK REVIEWS

Edited by W. C. H. PRENTICE, Wheaton College, Norton, Massachusetts

International Review of Research in Mental Retardation. Vol. 2. Edited by NORMAN R. ELLIS. New York: Academic Press, 1968. Pp. xi, 366. $11.50.

The appearance of this series of international reviews reflects the interest in and the volume of research on mental deficiency. In the last 10 years, several major books have been published, and increasing numbers of articles concerned with mental deficiency have appeared in several journals. Ellis has found the mental-deficiency literature sufficiently diversified, with respect both to subject matter and to geographical origin, to warrant a review series. Ellis' point here seems justified, whether the reviews are organized around an author's work or favorite approach (which seems to be the tendency in the present series) or are approached in survey as do the various Annual Reviews. Norman Ellis also has impressive credentials to edit this series. He has done extensive experimental work, has formulated the theory of impairment of short-term memory trace in the mentally defective, and is the editor of *Handbook in Mental Deficiency: Psychological Theory and Research* (1963).

If the first two volumes are typical, one might conclude that the title should have indicated that the concern is with psychological research. The first volume included many topics. The second volume seems to focus on learning problems. Six of its 10 chapters discuss learning and two are national reviews, while only two chapters deal with other topics. Even Herman Spitz—who has done most of his work in perception—has been pressed into the mold of a learning theorist. There is no indication whether a similar focus on learning is to be expected in future volumes or if other topics will receive special attention. To be sure, most of the activity in the field today is in learning, and this must be reflected in a review volume. It may, however, also be of real service to bring less publicized topics to the attention of the readers.

In the opening chapter, Denny presents the elicitation theory, reviews learning deficits of retardates, and describes an appartus for teaching concepts. Apparently as a result of insufficient data, the relationship between the theory and the training procedure is not clearly explained, and this chapter might better have been postponed until more data were available. The chapter by Herman Spitz contends that the forms which according to Gestalt principles are the easiest to perceive are also the easiest to learn. Problems of stimulus and response symmetry are discussed. Spitz regards organization of sensory input as one of the major functions of CNS. In a population where CNS impairment is frequent if not standard, the implications are far-reaching, because Spitz suggests that organization of material prior to presentation will improve the performance of the retardates. Rathe Karrer's chapter on autonomic nervous system functions is particularly timely because physiological functions

have, largely, been neglected in the study of mental deficiency. Karrer has appreciable skill in succinctly summarizing the pertinent points of an article and in organizing them so that trends become apparent. His is a good example of what a review article can accomplish. Sheppe and Turrisi discuss mediating-response theory and discriminative learning. It would seem that this chapter falls outside the scope of the series, since it is not research in mental retardation. The authors caution about making comparisons between normals and retardates, while pointing to the general usefulness of becoming familiar with discriminative learning problems. With such criteria for inclusion, the review series could become shapeless.

A review of research on learning sets and transfer of training by Kaufman and Prehm is a conscientious summary of 47 articles Learning sets and transfer of training are treated as completely separate. The authors complain that significant variation in experimental methods makes direct comparison among the studies difficult. It would, however, have been an advantage if the material could have been better integrated and the relevant trends discussed. Otherwise, the basis for a review is lacking and the reader has difficulty in orienting to such a compilation of findings. Sidman and Stoddard's chapter on programming perception and learning provides a very interesting account of how to determine the circle-ellipse difference threshold by nonverbal means. It is a history of the development of the authors' own teaching programs and is illustrated by case histories rather than by data and statistics. The chapter appears less technical than others and seems to address itself to a different audience. This reviewer would have preferred a much shorter version of the development and more emphasis on results obtained to date. In her chapter on programmed instruction techniques, Frances Greene has caught the first phase of this development. Hers is a competent piece of work, giving first a well-organized summary of the findings, following this with the extraction of the major trends in the material, and concluding with some sensible comments on future work. Greene has included material not yet published at the time of her writing and some projects which had just begun. In spite of this, the article will soon be outdated, if it has not already become so. This fact, rather than detracting from the chapter, only indicates that the continuation will soon have to be written.

Ivar Bjørgen's chapter takes on the discouraging task of reviewing research where no research exists. As he says, "thus far there have emerged no special traditions or results that seem to warrant considering separately the Norwegian research in this field" (p. 241). Consistent with this, no list of references is included. It might have been kinder to leave the Norwegians alone until there was something to report. Lafon and Chabanier write about recent research on mental deficiency in France. They describe philosophies and training programs and enumerate the areas where research has been conducted. While a general orientation to events in France is achieved, one is left with the feeling of superficiality because of such sketchiness as "Constantinides has investigated the effect of hereditary factors in the development of mental retardation" (p. 262). It would have been preferable to reduce the scope, so one could gain meaningful familiarity with some aspects of research efforts in France. The strong psychoanalytical attitude reflected in the article may surprise many American readers. (One wonders why references in this chapter do

not conform to those in other chapters.) Manny Sternlicht's chapter on psychotherapeutic procedures then covers 40 pages and 129 references. With some additions, it could well have been an independent publication. Sternlicht's treatment is refreshingly judicious, and the material is well organized. The question of adequate criteria for therapeutic success remains; the therapist's statement can not be considered sufficient. But this is a standard problem with psychotherapy. It is interesting to note that most of the work reported in this area has been done by psychologists; few psychiatrists have been active. In this respect there seems to be a difference between France and the United States.

Since so many people are involved in a series of reviews, it will take some time before the series finds its final form. Ellis' international review is well on its way. The idea of summary chapters from various foreign countries is a good one, but further thought should be given to their organization. The present volume has 598 references, and although not all of them are directly related to mental deficiency, they do indicate that such research is sufficiently extensive and sufficiently productive to warrant a review series. Ellis is making a significant contribution to the field with this work.

Institute for Basic Research in Mental Retardation, Staten Island, New York — JOHS. CLAUSEN

The Nature of Perceptual Adaptation. By IRVIN ROCK. New York: Basic Books, 1966. Pp. x, 289.

The author's basic intent is to bring under the wings of a general theory of space perception the effects and aftereffects of various kinds of transformations of the retinal image optically imposed by different sorts of lenses. Since he defines perception as "how the world appears," the writer declares himself a post-Watsonian thinker. Perceptual adaptation, in this context, means those forms of perceptual change that follow after the initial impression of distortion. Generally, it is directed to an approximation to 'veridicality,' despite the continuance of the physical source of distortion. Behavioral reorganization is not considered the primary factor within the process of adaptation.

After some remarks on the history of distortion experiments, techniques of measuring the amount of adaptation are discussed. In the subsequent chapters, the results of experimental inversion, reversal, tilting, displacement, magnification, and form distortion are mentioned. The description is combined with the presentation of several experimental results obtained by the author. In each of these items the reader will find a great deal of theoretical discussion designed to clarify the situation from the outset. A crucial question concerning inversion, reversal, and tilting, for instance, is how, in such cases, vision may be able to indicate any change at all. Assuming that the visual mechanism innately should be able to detect only changes in *relative* position, distance, and direction within the visual field, the *absolute* localization of the whole visual field can never be of a visual nature. Influences from other modalities, however, are supposed by the author to be strictly incompatible with vision. The solution of this puzzle fills the first third of the book (or even more) and reveals the roots of the author's theory: It is not touch, sound, vestibular and posture sensations, etc., that conflict with vision, but the visualized traces of

all such activities "previously formed (by association) between specific 'local signs' in each modality . . ." and, as a consequence, "it is the visualized location based on retinal local signs which must change for visual adaptation . . ." (p. 25). The author emphasizes an egocentric system of orientation (related to the sides of the head) which serves as a kind of anchor for 'uprightness' as well as 'leftness' or 'rightness,' especially in reduced-cue situations. From such considerations the author tends to support Charles Harris' theory of visual adaptation, *i.e.* that it consists rather in a change of the 'felt position' of the head, the arms, and legs, the whole situation thus tending to become 'harmonious' again despite the continuous wearing of the glasses. In connection with displacement experiments (by means of wedge prisms), a similar ambiguity of explanation is discussed. As a matter of fact, the apparent middle of the visual field (the median) becomes changed by wearing the prisms. The same is true for pointing with the hand. How should we decide whether the adaptive change is visual or proprioceptive? The reader will enjoy the creativity and subtility of the experiments presented to answer the question.

The role of active vs. passive behavior for visual adaptation is another focus of interest. The concepts of reafference, exafference, efference copy, corollary discharge, and others are introduced and explained. The result, as the author suspects (including adaptation effects under passive movements), confirms rather a general principle of coincidence between movement and visual change—a coincidence appearing to the author as more important than the voluntary aspect of movement (*i.e.* as self-produced vs. passive). In this connection the work of Held and his group is discussed. In connection with experiments on form distortion, the most important results and theoretical views are referred to and criticized, as are the adaptation to curvatures produced either by prisms or by looking to curved lines, the effects of half-prism experiments, and the question of gaze-contingent stimulation.

To sum up, Rock's book, at the present stage of discussion, belongs rather in the hands of specialists in perception than of students of general psychology. For the first group, the survey of data on perceptual adaptation is impressive and almost up to date. The theoretical considerations of the author are stimulating, even if the reader refuses the Helmholtzian-like background of the author's theory. Psychology students may become confused since it is not always easy to follow the author's reasoning.

Institut für Psychologie, Universität Innsbruck Y. Kohler

The Discrimination Process and Development. By Brian J. Fellows. Oxford: Pergamon Press, 1968. Pp. xii, 218. $10.00.

In a short and clearly-written book, four main topics are briefly discussed. They include discrimination-learning theory, developmental factors in discrimination learning, methods of teaching young children to discriminate differently oriented forms (principally mirror images), and perceptual theories relevant to form discrimination. Anyone wanting an introduction to these topics will find the book useful. The strength of the book lies in its eclectic approach to providing information that might be useful in teaching children to discriminate differently oriented forms. For example, theories of cognitive development and techniques of programmed instruction are both discussed. Although the

application of theoretical discussions and experimental findings to the teaching of form discrimination is not entirely convincing, there are enough plausible conclusions to merit the attention of educational psychologists involved in teaching children with reading problems, particularly reversal readers.

The book has several shortcomings, particularly when the author talks about discrimination learning. The theories are discussed very superficially, and extremely similar theoretical treatments are discussed in spaced chapters, making for poor organization. Also, the author's conclusions about the nature of the learning process are highly dubious. He seems to argue that animals and young children learn in an associative, S-R fashion, whereas older children and adults are likely to proceed "more thoughtfully." There is considerable evidence that opposes his conclusion about the manner in which young children and animals learn (*e.g.* experiments on intradimensional and extradimensional shifts, the overlearning reversal effect, and production vs. mediation deficiency) and, except for the latter case, this evidence was available before publication of the book. Also, when the author discusses an experiment conducted by Kurtz (pp. 25–27), he describes the subjects as young children rather than as the college students that they were and then contrasts the results of this experiment with other experiments which did involve young children. Furthermore, the treatment of Kurtz's theoretical position is inaccurate. But even though the book has several deficiencies, it would still seem useful as a supplementary text in a course on learning in children, cognitive development, or discrimination learning—except for its unreasonably high cost.

University of Washington F. MICHAEL RABINOWITZ
JOSEPH C. CAMPIONE

To Change a Child: A Report on the Institute for Developmental Studies. By FRED POWLEDGE. Chicago: Quadrangle Books, 1968. Pp. vii, 110. Cloth, $5.50; paperback, $2.25.

Fred Powledge has written, simply and cogently, an overview of work done by the Institute for Developmental Studies of New York University in restructuring the early school experience of the young disadvantaged lower-class child. It is the Institute's working premise that the development of a child's cognitive skills and the improvement of his language and of his perceptual and conceptual abilities in early schooling can change not only his academic and occupational expectancies but also his self-concept. 'Take care of the cognitive deficits of the disadvantaged child and he will come out with fewer emotional scars.'

The Institute's functions are at least tri-pronged and each function helps to feed the others. It conducts research on a continuing basis and both demonstrates and evaluates an enrichment program for disadvantaged children from preschool through third grade. All school situations are analyzed to the end of maximizing learning; the first greeting, the casual conversations between pupil and teacher, the snack period, the response to rhythm, are all studied. Nothing is taken for granted, whether that the child knows what a living room or dining room is or how many legs a horse has, that he is tuned in to what the teacher is saying, or that he responds to the school's reward system. 'Is the questioning too simple or too complex?' 'In what way shall the teacher

introduce letters to the child?' 'How can such devices as the telephone be used to best advantage?'

The Institute believes, according to Powledge, that a program of massive intervention of the kind they suggest cannot be accomplished within the framework of education as it is today. The book hints at various requirements (including changes in the bureaucratic structure, in teacher training, in student-teacher ratios) without providing any analysis of an overall reconstruction. But perhaps this is the subject for another book.

LILLIAN COHEN KOVAR

BOOKS RECEIVED

(The books listed here have not as yet been noted in our pages. Listing here does not, however, preclude their later review.)

Allison, Joel, Blatt, Sidney J., and Zimet, Carl N. *The Interpretation of Psychological Tests*. New York: Harper & Row, 1968. Pp. x, 342. $8.75.

Amos, William E., and Grambs, Jean Dresden, eds. *Counseling the Disadvantaged Youth*. Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. x, 438. $7.95.

Anastasi, Anne. *Psychological Testing*. 3rd ed. New York: Macmillan, 1968. Pp. viii, 665. $9.25.

Andreas, Burton G. *Psychological Science and the Educational Enterprise*. New York: Wiley, 1968. Pp. ix, 103. Cloth, $6.95; paperback, $3.95.

Argyle, Michael. *Psychology and Social Problems*. New York: Barnes & Noble, 1968. Pp. 232. Cloth, $4.00; paperback, $2.75.

Auerbach, Alice B. *Parents Learn through Discussion: Principles and Practices of Parent Group Education*. New York: Wiley, 1968. Pp. xii, 358. $7.95.

Ausubel, David P. *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart and Winston, 1968. Pp. xvii, 685. $8.95.

Bakan, David. *The Duality of Human Existence*. Chicago: Rand McNally, 1966. Pp. 242. $6.50.

Baldwin, Alfred L. *Theories of Child Development*. New York: Wiley, 1967. Pp. xii, 618. $8.95.

Bannister, D., and Mair, J. M. M. *The Evaluation of Personal Constructs*. London: Academic Press, 1968. Pp. 232. 50*s*./$8.00.

Barlow, John A. *Stimulus and Response*. New York: Harper & Row, 1968. Pp. xviii, 199. Paperback.

Barnes, Charles D., and Kircher, Christopher, eds. *Readings in Neurophysiology*. New York: Wiley, 1968. Pp. viii, 482. $9.50.

Bateman, Barbara D. *Interpretation of the 1961 Illinois Test of Psycholinguistic Abilities*. Seattle: Special Child Publications, 1968. Pp. 108. Paperback, $3.00.

Bauer, Raymond A., and Gergen, Kenneth J. *The Study of Policy Formation*. New York: The Free Press, 1968. Pp. xxii, 392. $9.95.

Bauer, Raymond, and Greyser, Stephen A. *Advertising in America: The Consumer View*. Boston: Division of Research, Harvard Business School, 1968. Cloth, $9.00; paperback, $5.00.

Beam, Lura. *He Called Them by the Lightning*. Indianapolis: Bobbs-Merrill, 1967. Pp. 230. $5.75.

Becker, P. E. *Humangetik*. Germany: Georg Thieme Verlag, 1968. Pp. xv, 535. $45.50.

Beech, H. R., and Fransella, Fay. *Research and Experiment in Stuttering*. New York: Pergamon Press, 1968. Pp. xi, 224. $9.00.

Beecroft, Robert S. *Classical Conditioning.* Goleta, Calif.: Psychonomic Press, 1966. Pp. ii, 198.

Beradt, Charlotte. *The Third Reich of Dreams.* Chicago: Quadrangle Books, 1968. Pp. 177. $4.95.

Bigge, Morris L., and Hunt, Maurice P. *Psychological Foundations of Education.* 2nd ed. New York: Harper & Row, 1968. Pp. xiii, 530. $9.95.

Blanck, Rubin, and Blanck, Gertrude. *Marriage and Personal Development.* New York: Columbia University Press, 1968. Pp. xiv, 191. $6.00.

Blum, Milton L., and Naylor, James C. *Industrial Psychology: Its Theoretical and Social Foundations.* New York: Harper & Row, 1968. Pp. xiii, 633. $9.95.

Bock, R. Darrell, and Jones, Lyle V. *The Measurement and Prediction of Judgement and Choice.* San Francisco: Holden-Day, 1968. Pp. x, 370. $13.75.

Bohm, Ewald. *Psychodiagnostisches Vademecum.* Bern: Hans Huber, 1967. Pp. 198.

Bonham, Marilyn. *The Laughter and Tears of Children.* New York: Macmillan, 1968. Pp. 213. $4.95.

Brammer, Lawrence M., and Shostrom, Everett L. *Therapeutic Psychology: Fundamentals of Actualization Counseling and Psychotherapy.* 2nd ed. Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. ix, 486. $8.95.

Branch, C. H. *Aspects of Anxiety.* 2nd and enlarged ed. Philadelphia: Lippincott, 1968. Pp. 148. $4.50.

Brown, Robert. *Explanation in Social Science.* Chicago: Aldine, 1963. Pp. viii, 198. Cloth, $6.50; paperback, $2.95.

Buckley, Walter. *Modern Systems Research for the Behavioral Scientist: A Sourcebook for the Application of General Systems Theory to the Study of Human Behavior.* Chicago: Aldine, 1968. Pp. xxv, 525. $14.75.

Bures, Jan, Petran, Mojmir, and Zachar, Jozef, eds. *Electrophysiological Methods in Biological Research.* New York: Academic Press, 1967. Pp. 824. $28.00.

Carlozzi, Carl G. *Death and Contemporary Man: The Crisis of Terminal Illness.* Grand Rapids, Mich.: Wm. B. Eerdmans, 1968. Pp. 79. Paperback, $1.45.

Cartwright, Dorwin, and Zander, Alvin. *Group Dynamics: Research and Theory.* 3rd ed. New York: Harper & Row, 1968. Pp. ix, 580. $11.95.

Chaplin, James P., and Drawiec, T. S. *Systems and Theories of Psychology.* New York: Holt, Rinehart and Winston, 1968. Pp. xiv, 624. $8.95.

Chapman, A. H. *Put Offs and Come Ons.* New York: Putnam, 1968. Pp. 258. $5.95.

Clark, Virginia, and Tarter, Michael E. *Preparation for Basic Statistics.* New York: McGraw-Hill, 1968. Pp. x, 150. $3.95.

Cohen, Hyman L., and Brumlik, Joel. *A Manual of Electroneuromyography.* New York: Hoeber, 1968. Pp. 192. $10.50.

Coppersmith, Stanley. *The Antecedents of Self-esteem.* San Francisco: W. H. Freeman, 1967. Pp. ix, 283. $6.00.

Cope, Oliver. *Man, Mind, and Medicine.* Philadelphia: Lippincott, 1968. Pp. 144. Cloth, $4.25; paperback, $1.95.

Corso, John F. *The Experimental Psychology of Sensory Behavior.* New York: Holt, Rinehart and Winston, 1967. Pp. xii, 628. $12.95.

Crosby, Andrew. *Creativity and Performance in Industrial Organization.* New York: Barnes & Noble, 1968. Pp. xii, 197. $6.75.

Cumming, Elaine. *Systems of Social Regulation.* New York: Atherton Press, 1968. Pp. vii, 324. $7.95.

De Becker, Raymond. *The Understanding of Dreams.* New York: Hawthorn Books, 1968. Pp. xiii, 432. $7.95.

De Bold, Richard C. *Manual of Contemporary Experiments of Psychology.* Englewood Cliffs, N. J.: Prentice-Hall. 1968. Pp. 245.

De Cecco, P. John. *The Psychology of Learning and Instruction: Educational Psychology.* Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. xvi, 800. $9.25.

Deutsch, M., Katz, I., and Jensen, A. R., eds. *Social Class, Race, and Psychological Development.* New York: Holt, Rinehart and Winston, 1968. Pp. 423. $7.95.

Dewey, John. *Psychology.* (*The Early Works of John Dewey, 1882–1898,* Vol. 2.) Carbondale: Southern Illinois University Press, 1967. Pp. cix, 366. $10.00.

De Zwart, H. Sinclair. *Acquisition du langage et développement de la pensée.* Paris: Dunod, 1967. Pp. 168. Paperback.

Dickson, William J., and Roethlisberger, F. J. *Counseling in an Organization: A Sequel to the Hawthorne Researches.* Boston: Division of Research, Graduate School of Business Administration, Harvard University, 1966. Pp. xviii, 480. $10.00.

Donahue, George T., and Nichtern, Sol. *Teaching the Troubled Child.* New York: The Free Press, 1968. Pp. xvi, 206. Paperback, $2.45.

Dreikurs, Rudolf. *Psychology in the Classroom: A Manual for Teachers.* 2nd ed. New York: Harper & Row, 1968. Pp. xvi, 284. Cloth, $5.95; paperback, $3.75.

Dreikurs, Rudolf, and Gray, Loren. *Logical Consequences: A New Approach to Discipline.* New York: Meredith Press, 1968. Pp. 216. $4.95.

Dyal, James A. *Readings in Psychology: Understanding Human Behavior.* 2nd ed. New York: McGraw-Hill, 1967. Pp. xi, 500. Cloth, $6.50; paperback, $4.50.

Edwards, Allen L. *Experimental Design in Psychological Research.* 3rd ed. New York: Holt, Rinehart and Winston, 1968. Pp. xix, 455. $8.95.

Edwards, David C. *General Psychology.* New York: Macmillan, 1968. Pp. xiii, 394. $5.95.

Ellis, Norman R., ed. *Research in Mental Retardation.* Vol. 2. New York: Academic Press, 1966. Pp. xii, 366. $11.50.

Endler, Norman S., Boulter, Lawrence R., and Osser, Harry. *Contemporary Issues in Developmental Psychology.* New York: Holt, Rinehart and Winston, 1968. Pp. xvii, 682. $9.95.

Engel, J., Kollat, D., and Blackwell, R. *Consumer Behavior.* New York: Holt, Rinehart and Winston, 1968. Pp. 652. $12.95.

Etzioni, Amitai. *The Active Society: A Theory of Societal and Political Processes.* New York: The Free Press, 1968. Pp. xxv, 698. $10.95.

Evans, Ellis D., ed. *Children: Readings in Behavior and Development.* New York: Holt, Rinehart and Winston, 1968. Pp. xv, 571. Cloth, $8.95; paperback, $5.95.

Fairweather, George W. *Methods for Experimental Social Innovation.* New York: Wiley, 1967. Pp. x, 250. $7.95.

Fantini, Mario, and Weinstein, Gerald. *Making Urban Schools Work: Social*

Realities and the Urban School. New York: Holt, Rinehart and Winston, 1968. Pp. x, 62. Paperback, $1.50.

Fass, Jerome. *A Primer for Parents.* New York: Trident Press, 1968. Pp. 192. $4.95.

Fellows, Brian J. *The Discrimination Process and Development.* Oxford: Pergamon Press, 1968. Pp. xii, 218. $10.00.

Ferman, Louis A., Kornbluh, Joyce L., and Haber, Alan, eds. *Poverty in America.* Ann Arbor: University of Michigan Press, 1968. Pp. xxxii, 669. Cloth, $9.75; paperback, $5.75.

Finkelstein, Sidney. *Sense and Nonsense of Marshall McLuhan.* New York: International Publishers, 1968. Pp. 128. Cloth, $4.95; paperback, $1.45.

Flavell, John H. *The Development of Role-taking and Communication Skills in Children.* New York: Wiley, 1968. Pp. vii, 239. $9.50.

Frances, Robert. *Psychologie de l'esthetique.* Paris: Presses Universitaires de France, 1968. Pp. 204. Paperback.

Frank, Jerome D. *Sanity and Survival: Psychological Aspects of War and Peace.* New York: Random House, 1967. Pp. x, 330. $5.95.

Frost, Joe L. *Early Childhood Education Rediscovered: Readings.* New York: Holt, Rinehart and Winston, 1968. Pp. xiv, 594. Paperback, $5.95.

Gardner, Riley W., and Moriarty, Alice. *Personality Development at Preadolescence.* Seattle: University of Washington Press, 1968. Pp. ix, 344. $9.50.

Gellhorn, Ernst. *Autonomic-Somatic Integrations: Physiological Basis and Psychological and Clinical Implications.* Minneapolis: University of Minnesota Press, 1967. Pp. xiii, 318. $10.00.

Gesell, Arnold, and others. *Vision: Its Development in Infant and Child.* New York: Hafner, 1967. Pp. xv, 329. $11.50.

Gnagey, William J. *The Psychology of Discipline in the Classroom.* New York: Macmillan, 1968. Pp. xi, 81. Paperback, $1.50.

Goffman, Erving. *Interaction Ritual.* Chicago: Aldine, 1968. Pp. 270. $5.75.

Gordon, C., and Gergen, K. J., eds. *The Self in Social Interaction.* Vol. 1. New York: Wiley, 1968. $9.95.

Gordon, William J. J. *Synectics: The Development of Creative Capacity.* New York: Collier, 1968. Pp. x, 180. Paperback, $1.25.

Gorlow, Leon, and Katkovsky, Walter. *Readings in the Psychology of Adjustment.* 2nd ed. New York: McGraw-Hill, 1968. Pp. xiii, 586. Cloth, $6.95; paperback, $3.95.

Grant, Vernon W. *Great Abnormals.* New York: Hawthorn, 1968. Pp. 248. $5.95.

Gribbons, Warren D., and Lohnes, Paul R. *Emerging Careers.* New York: Teachers College, Columbia University, 1968. Pp. v, 202. $5.95.

Groffmann, K. J., and Wewetzer, K. H. *Person Als Prozess.* Bern and Stuttgart: Verlag Hans Huber, 1968. Pp. 242.

Gronlund, Norman E. *Readings in Measurement and Evaluation.* New York: Macmillan, 1968. Pp. xii, 463. Paperback, $4.95.

Grossman, Editha Maria. *Die Problematik des Dazwischen-stehens.* (Monographien und Studien zur Konfliktpsychologie, Abt. I, Bd. 5.) Munich and Basel: Ernst Reinhardt Verlag, 1967. Pp. 176.

Haber, Ralph Norman, ed. *Contemporary Theory and Research in Visual Perception.* New York: Holt, Rinehart and Winston, 1968. Pp. xi, 814. $14.95.

Happenie, Sylvene Dillon. *Colostomy: A Second Chance.* Springfield, Ill.: Charles C Thomas, 1968. Pp. viii. 86. $4.75.

Hartmann, Ernest, and others. *Adolescents in a Mental Hospital.* New York: Grune & Stratton. 1968. Pp. ix, 197. Paperback. $5.75.

Heimler, Eugene. *Mental Illness and Social Work.* Baltimore, Md.: Penguin Books. 1968. Pp. 160. Paperback. $1.25.

Herbst, P. G. *Autonomous Group Functioning: An Exploration in Behavior Theory and Measurement.* New York: Barnes & Noble, 1968. Pp. xiii, 271.

Hess, Robert D., and Bear, Roberta Meyer. *Early Education: A Comprehensive Evaluation of Current Theory, Research, and Practice.* Chicago: Aldine, 1968. Pp. x, 272. $6.95.

Hill, Brian. *Gates of Horn and Ivory: An Anthology of Dreams.* New York: Taplinger, 1968. Pp. xix, 216. $5.95.

Hoffman, Everett. *The Coming Up of Mr. Rattus: A Fable.* Chicago: Adams Press, 1968. Pp. 105. Paperback, $2.95.

Hollister, Leo E. *Chemical Psychoses: LSD and Related Drugs.* Springfield, Ill.: Charles C Thomas. 1967. Pp. xi. 190. $8 00.

Huey, Edmund Burke. *The Psychology and Pedagogy of Reading.* Cambridge, Mass.: The M. I. T. Press, 1968. Pp. xlix, 469. Paperback, $3.95.

The Human Condition: Lectures Delivered during the University of New Hampshire's Centennial Year. Vol. I, *The Nature of New England.* Vol. II, *Man in Science.* Vol. III, *Man in Society.* Durham: University of New Hampshire Press, 1967.

Hunt, James. *Stammering and Stuttering: Their Nature and Treatment.* Originally published in 1870. New York: Hafner. 1967. Pp. 182. $5.50.

Hyman, Herbert H., and Singer, Eleanor. *Readings in Reference Group Theory and Research.* New York: The Free Press, 1968. Pp. xi. 509. $10.95.

Jersild, Arthur T. *Child Psychology.* 6th ed. Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. xii, 610. $8.50.

Jessor, Richard, and others. *Society, Personality, and Deviant Behavior: A Study of a Tri-Ethnic Community.* New York: Holt, Rinehart and Winston, 1968. Pp. xi, 500.

Johnson, Marjorie Seddon, and Kress, R. A., eds. *Significant Issues in Reading: Proceedings of the 23rd Annual Reading Institute at Temple University.* Philadelphia: Temple University Reading Clinic, 1968. Pp. 119. Paperback, $3.00.

Jones, Marshall R., ed. *Miami Symposium on the Prediction of Behavior, 1967: Aversive Stimulation.* Coral Gables: University of Miami Press, 1968. Pp. 145. $5.95.

Juurmaa, Jyrki. *Ability Structure and Loss of Vision.* New York: American Foundation for the Blind, 1968. Pp. 128. Paperback, $2.00.

Kagan, Jerome, and Havemann, Ernest. *Psychology and Introduction.* New York: Harcourt, Brace & World, 1968. Pp. 673. $8.25.

Katona, George. *Organizing and Memorizing: Studies in the Psychology of Learning and Teaching.* New York: Hafner, 1940. Pp. xii, 318. $9.00.

Kaufmann, Pierre. *Kurt Lewin: Une theorie du champ dans les sciences de l'homme.* France: Librarie Philosophique (J. Vrin). Pp. 381.

Kellogg, W. N., and Kellogg, L. A. *The Ape and the Child.* New York: Hafner, 1933. Pp. xiv, 341. $10.50.

Kells, Lyman M. *Differential Equations: A Brief Course with Applications.* New York: McGraw-Hill, 1968. Pp. xi, 252. Paperback, $7.50.

Kenshalo, Dan, ed. *The Skin Senses.* Springfield, Ill.: Charles C Thomas. Pp. xvii, 636. $26.00.

Kiev, Ari. *Curanderismo: Mexican-American Folk Psychiatry.* New York: The Free Press. 1968. Pp. xiii, 207. $6.95.

Kirkpatrick, James J., Ewen, Robert B., Barrett, Richard S., and Katzell, Raymond A. *Testing and Fair Employment: Fairness and Validity of Personnel Tests for Different Ethnic Groups.* New York: New York University Press, 1968. Pp. x. 145. $5.00.

Knight, James A. *For the Love of Money.* Philadelphia: Lippincott, 1968. Pp. 184. $4.95.

Kockelmans, Joseph J. *A First Introduction to Husserl's Phenomenology.* Pittsburgh: Duquesne University Press, 1968. Pp. xxiii, 372. $8.95.

Koestler, Arthur. *The Ghost in the Machine.* New York: Macmillan, 1968. Pp. xiv, 384. $6.75.

Kolers, Paul A., and Eden, Murray, eds. *Recognizing Patterns: Studies in Living and Automatic Systems.* Cambridge, Mass.: The M. I. T. Press, 1968. Pp. xi, 237. $11.00.

Koppitz, Elizabeth Munsterberg. *Psychological Evaluation of Children's Human Figure Drawings.* New York: Grune & Stratton, 1968. Pp. x, 341. $9.75.

Kovar, Lillian Cohen. *Faces of the Adolescent Girl.* Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. xi, 178. $4.95.

Kreitler, Hans, and Kreitler, Shulamith. *Die Kognitive Orientierung des Kindes.* Munich and Basel: Ernst Reinhardt Verlag, 1967. Pp. 205.

Kummer, Hans. *Social Organization of Hamadryas Baboons: A Field Study.* Chicago: University of Chicago Press, 1968. Pp. 189. $8.95.

Kusch, Lothar, and Rosenthal, H. J. *Mathematik.* Essen: Verlag W. Girardet, 1967. Pp. 314. Paperback.

Langendoen, D. Terence. *The London School of Linguistics: A Study of the Linguistic Theories of B. Malinowski and J. R. Firth.* Cambridge, Mass.: The M. I. T. Press. Pp. xii, 123. $5.95.

Levine, David. ed. *Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press, 1967. Pp. ix, 335. Cloth, $6.25; paperback, $3.25.

Levy, Sheldon G. *Inferential Statistics in the Behavioral Sciences.* New York: Holt, Rinehart and Winston, 1968. Pp. vii, 267. $7.95.

Lichter, Solomon O., Rapien, Elsie B., Siebert, Frances M., and Sklansky, Morris A. *The Drop-Outs.* New York: The Free Press, 1968. Pp. xiii, 302. Paperback, $2.95.

Lindesmith, Alfred R. *Addiction and Opiates.* Chicago: Aldine, 1967. Pp. vii, 295. $7.50.

Lindgren, Henry Clay. *Readings in Educational Psychology.* New York: Wiley, 1968. Pp. x, 474. Cloth, $8.95; paperback, $5.95.

Lindsley, Donald B., and Lumsdaine, Arthur A., eds., *Brain Function and Learning. (Brain Function,* Vol. IV.) Berkeley: University of California Press, 1968. Pp. xiii, 364. $15.00.

London, Perry, and Rosenhan, David. *Foundations of Abnormal Psychology.* New York: Holt, Rinehart and Winston, 1968. Pp. x, 644. $9.95.

Louria, Donald B. *The Drug Scene.* New York: McGraw-Hill, 1968. Pp. viii, 215. $5.95.

Lovejoy, Elijah. *Attention in Discrimination Learning*. San Francisco: Holden-Day, 1968. Pp. xii, 133. $8.50.

McDavid, John W., and Harari, Herbert. *Social Psychology: Individuals, Groups, Societies*. New York: Harper & Row, 1968. Pp. xi, 479.

McDougall, William. *Explorer of the Mind: Studies in Psychical Research*. New York: Garrett Publications, 1968. Pp. 319. $8.50.

McGregor, Douglas. *Leadership and Motivation*. Cambridge, Mass.: The M. I. T. Press, 1968. Pp. xxiii, 286. Paperback.

McGuigan, F. J. *Experimental Psychology: A Methodological Approach*. 2nd ed. Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. xii, 400. $8.50.

MacKay, R. Stuart. *Bio-Medical Telemetry*. New York: Wiley, 1968. Pp. xi, 388. $12.50.

Mallison, Ruth. *Education as Therapy: Suggestions for Work with Neurologically Impaired Children*. Seattle: Special Child Publications, 1968. Pp. 166. Paperback, $3.50.

Manuel, Frank E. *A Portrait of Isaac Newton*. Cambridge, Mass.: The Belknap Press, 1968. $11.95.

Mechanic, David. *Medical Sociology: A Selective View*. New York: Macmillan, 1968. Pp. 504.

Mehrabian, Albert. *An Analysis of Personality Theories*. Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. xii, 217.

Meili, R., and Rohracher, H. *Lehrbuch der experimentellen Psychologie*. Bern and Stuttgart: Verlag Hans Huber, 1968. Pp. xii, 501.

Merton, Robert K. *Social Theory and Social Structure*. New York: The Free Press, 1968. Pp. xxiii, 702. $8.95.

Mitchell, John H. *Writing for Professional and Technical Journals*. New York: Wiley, 1968. Pp. 405. $7.50.

Moeller, Helen. *Tornado*. Westwood, N. J.: Fleming H. Revell Co., 1967. Pp. 109. $3.50.

Moles, Azraham, tr. by Joel E. Cohen. *Information Theory and Esthetic Perception*. Urbana: University of Illinois Press, 1968. Pp. 217. Cloth, $7.50; paperback, $2.45.

Moncreiff, William P. *Refraction: Neurophysiological and Psychological Viewpoints*. Springfield, Ill.: Charles C Thomas, 1968. Pp. xv, 92. $5.75.

Montmollin, Maurice de. *Les Systemes hommes-machines*. Paris: Presses Universitaires de France, 1967. Pp. 248. Paperback.

Mouly, J. George. *Psychology for Effective Teaching*. 2nd ed. New York: Holt, Rinehart and Winston, 1968. Pp. xii, 639. $8.95.

Moustakas, Clark. *Creativity and Conformity*. (Insight No. 40.) Princeton, N. J.: Van Nostrand, 1967. Pp. xv, 142. Paperback, $1.95.

Mowrer, O. Hobart, ed. *Morality and Mental Health*. Chicago: Rand McNally, 1967. Pp. xvii, 669. $9.00.

Murphy, Gardner, and Spohn, Herbert E. *Encounter with Reality: New Forms for an Old Quest*. Boston: Houghton Mifflin, 1968. Pp. xiii, 162. $6.50.

Myers, Jerome K., and Bean, Lee L. *A Decade Later*. New York: Wiley, 1968. Pp. ix, 250. $7.95.

Myrdal, Alva. *Nation and Family*. Cambridge, Mass.: The M.I.T. Press, 1968. Pp. xxxiii, 441. Paperback, $3.45.

Newson, John, and Newson, Elizabeth. *Four Years Old in an Urban Community*. Chicago: Aldine, 1968. Pp. 570. $9.75.

Newton, Grant, and Levine, Seymour. *Early Experience and Behavior: The Psychobiology of Development*. Springfield, Ill.: Charles C Thomas, 1968. Pp. xii, 785. $28.50.

Noll, Victor H., and Noll, Rachel P. *Readings in Education Psychology*. 2nd ed. New York: Macmillan, 1968. Pp. xii, 432. Paperback, $4.50.

Osborn, Frederick, introd. by T. Dobzhansky. *The Future of Human Heredity: An Introduction to Eugenics in Modern Society*. New York: Weybright and Talley, 1968. Pp. vii, 133. $5.95.

Packard, Vance. *The Sexual Wilderness*. New York: David McKay, 1968. Pp. xi, 553. $6.95.

Parmenter, Ross. *The Awakened Eye*. Middletown, Conn.: Wesleyan University Press, 1968. Pp. xii, 352. $8.50.

Petras, John W. *George Herbert Mead: Essays on His Social Philosophy*. New York: Teachers College, Columbia University. Pp. ix, 161. $7.50.

Plutchik, Robert. *Foundations of Experimental Research*. New York: Harper & Row, 1967. Pp. xiii, 290. Paperback, $3.95.

Postlethwaite, Neville. *School Organization and Student Achievement: A Study Based on Achievement in Mathematics in Twelve Countries*. New York: Wiley, 1968. Pp. 146. $8.00.

Pranger, Robert J. *Action, Symbolism, and Order: The Existential Dimensions of Politics in Modern Citizenship*. Knoxville, Tenn.: Vanderbilt University Press, 1968. Pp. ix, 225. $6.95.

Pye, Lucian. *The Spirit of Chinese Politics: A Psychocultural Study of the Authority Crisis in Political Development*. Cambridge, Mass.: The M. I. T. Press, 1968. Pp. xxii, 255. $8.95.

Rabin, A. I. *Projective Techniques in Personality Assessment*. New York: Springer, 1968. Pp. x, 638.

Ramwell, Peter W., and Shaw, Jane E., eds. *Prostaglandin: Symposium of the Worcester Foundation for Experimental Biology*. New York: Wiley, 1968. Pp. xxiv, 402. $15.00.

Rapoport, Robert N., with Rhoda Rapoport and Irving Rosow. *Community as Doctor*. New York: Barnes & Noble, 1968. Pp. x, 325. Cloth, $7.25; Paperback, $4.25.

Revoredo, Emilio Barrantes. *Educacion*. Lima: Universidad Nacional Mayor de San Marcos, 1965. Pp. 311. Paperback.

Richardson, Elizabeth. *The Environment of Learning*. New York: Weybright and Talley, 1967. Pp. viii, 247. $6.00.

Rosenbaum, Peter S. *The Grammar of English Predicate Complement Constructions*. Cambridge, Mass.: The M. I. T. Press, 1967. Pp. xii, 128. $5.00.

Rosenberg, Marshall B. *Diagnostic Teaching*. Seattle: Special Child Publications, 1968. Pp. 125. Paperback, $3.50.

Rosenberg, Sheldon, and Koplin, James H. *Developments in Applied Psycholinguistics Research*. New York: Macmillan, 1968. Pp. vii, 311. $8.95.

Rosenthal, Robert, and Jacobson, Lenore. *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*. New York: Holt, Rinehart and Winston, 1968. Pp. xi, 240. Paperback, $3.95.

Rushing, William A., ed. *Deviant Behavior and Social Progress*. Chicago: Rand McNally. Pp. ix, 427. $4.50.

Schane, Sanford A. *French Phonology and Morphology*. Cambridge, Mass.: The M. I. T. Press, 1968. Pp. xx, 161. $7.50.

Scheffler, Israel. *Science and Subjectivity*. Indianapolis: Bobbs-Merrill, 1967. Pp. vi, 132. Paperback, $2.25.

Schestakow, A. W. *L'Enseignement programme et les machines a enseigner en U. R. S. S.* Paris: Dunod. Pp. 210.

Schneider, Daniel E. *Psychoanalysis of Heart Attack*. New York: Dial Press, 1967. Pp. xxvii, 234. $6.00.

Schoenberg, Bernard, Pettit, Helen F., and Carr, Arthur C., eds. *Teaching Psychosocial Aspects of Patient Care*. New York: Columbia University Press, 1968. Pp. xii, 420. $8.50.

Sem-Jacobsen, C. W. *Depth-Electrographic Stimulation of the Human Brain and Behavior*. Springfield, Ill.: Charles C Thomas, 1968. Pp. xv, 222. $15.50.

Senn, J. E., and Hartford, Claire, eds. *The Firstborn: Experiences of Eight American Families*. Cambridge, Mass.: Harvard University Press. Pp. 533. $10.00.

Sherif, Muzafer. *Social Interaction: Process and Products*. Chicago: Aldine, 1968. Pp. vii, 512. $9.75.

Siegel, Max. *The Counseling of College Students*. New York: The Free Press, 1968. Pp. xx, 467. $9.95.

Sigel, Irving E., and Hooper, Frank H., eds. *Logical Thinking in Children*. New York: Holt, Rinehart and Winston, 1968. Pp. xv, 541. $7.95.

Simmel, Marianne L. *The Reach of Mind: Essays in Memory of Kurt Goldstein*. New York: Springer, 1968. Pp. xii, 300. $9.00.

Singer, G., Bennett, Adrienne, and Day, R. H. *Perception: A Laboratory Manual*. N. S. W.: New South Wales University Press, 1967. Pp. ii, 151. Paperback.

Sites, Paul. *Lee Harvey Oswald and the American Dream*. New York: Pageant Press, 1967. Pp. 261. $4.00.

Skinner, B. F. *The Technology of Teaching*. New York: Appleton-Century-Crofts, 1968. Pp. ix, 271. Paperback, $2.95.

Smilansky, S. *The Effects of Sociodramatic Play on Disadvantaged Pre-School Children*. New York: Wiley, 1968. Pp. 164. $7.50.

Snyder, H. I. *Contemporary Educational Psychology*. New York: Wiley, 1968. Pp. xii, 236. $8.95.

Soddy, Kenneth, with Mary C. Kidson. *Men in Middle Life*. Philadelphia: Lippincott, 1967. Pp. xii, 485. $12.50.

Somjee, A. H. *Political Theory of John Dewey*. New York: Teachers College Press, 1968. Pp. xv, 208. $7.50.

Spitzer, Stephan P., and Denzin, Norman K. *The Mental Patient: Studies in the Sociology of Deviance*. New York: McGraw-Hill, 1968. Pp. xiv, 486. Cloth, $8.50; paperback, $5.95.

Staats, Arthur W. *Learning, Language, and Cognition: Theory, Research, and Methods for the Study of Human Behavior and Its Development*. New York: Holt, Rinehart and Winston, 1968. Pp. xvii, 614. $9.50.

Steidl, Rose E., and Bratton, Esther Crew. *Work in the Home*. New York: Wiley, 1968. Pp. xvi, 419. $9.95.

Stein, Calvert. *Practical Psychotherapeutic Techniques*. Springfield, Ill.: Charles C Thomas, 1967. Pp. xiii, 201. $7.50.

Stekel, Wilhelm. *The Interpretation of Dreams.* New York: Washington Square Press, 1967. Pp. x, 502. Paperback, $1.45.

Stenhouse, Lawrence. *Culture and Education.* New York: Weybright and Talley, 1967. Pp. viii, 156. $5.00.

Stephenson, William. *The Play Theory of Mass Communication.* Chicago: University of Chicago Press, 1967. Pp. x. 225. $5.00.

Stolz, Lois Meek. *Influences on Parent Behavior.* Stanford. Calif.: Stanford University Press, 1967. Pp. viii, 355. $8.95.

Studer-Salzmann. *Krimialitat, Erziehung und Ethik.* Bern: Verlag Hans Huber, 1967. Pp. 354.

Szondi, L. *Freiheit und Zwang im Schicksal des Einzelnen.* Bern: Verlag Hans Huber, 1968. Pp. 96.

Thompson, Travis, and Schuster, Charles R. *Behavioral Pharmacology.* Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. xvi, 297. $10 50.

Thurmond, Margaret E. *Neurotic Trends of Sir Willoughby: An Analysis of George Meredith's Characterization in "The Egoist."* California: Lorrin L. Morrison. 1968. Pp. 30. Paperback, $1.75.

Trabasso, Tom, and Bower, Gordon H. *Attention in Learning: Theory and Research.* New York: Wiley, 1968. Pp. xi, 253. $7.95.

Traxel, Werner. *Ueber Gegenstand und Methode der Psychologie.* Bern and Stuttgart: Verlag Hans Huber, 1968. Pp. 122. Paperback.

Turner, F. Bernadette. *God-Centered Therapy.* New York: Robert Speller & Sons, 1968. Pp. xvii, 277. $3.95.

Uttal, William R. *Real-Time Computers: Technique and Applications in the Psychological Sciences.* New York, Evanston, and London: Harper & Row, 1968. Pp. xiv. 338.

Vail, David J. *Mental Health Systems in Scandinavia.* Springfield, Ill.: Charles C Thomas, 1968. Pp. xxvii, 157. $8.00.

Valentine, C. W. *The Experimental Psychology of Beauty.* New York: Barnes & Noble, 1968. Pp. xi, 438. Paperback, $4.25.

Von Bertalanffy, Ludwig. *Robots, Men, and Minds: Psychology in the Modern World.* New York: George Braziller, 1967. Pp. x, 150. $5.00.

Von Lindholm, Gosta Olof. *Fehlerquellen der Sog. ASW-Versuche.* Helsinki: Academia Scientiarum Fennica, 1967. Pp. 232. Paperback.

Walker, Sydney, III. *Psychiatric Signs and Symptoms Due to Medical Problems.* Springfield, Ill.: Charles C Thomas, 1967. Pp. xxi, 299. $11.50.

Wallace, Karl R. *Francis Bacon on the Nature of Man.* Urbana: University of Illinois Press, 1967. Pp. 200. $5.75.

Warren, Richard M., and Warren, Roslyn P. *Helmholtz on Perception: Its Physiology and Development.* New York: Wiley, 1968. Pp. x, 277. $9.95.

Wathen-Dunn, Weiant. *Models for the Perception of Speech and Visual Form.* Cambridge, Mass.: The M. I. T. Press. Pp. x, 470. $10.00.

Weiskrantz, Lawrence. *Analysis of Behavioral Change.* New York: Harper & Row, 1968. Pp. 447.

Weisman, H. M. *Technical Correspondence: A Handbook and Reference Source for the Technical Professional.* New York: Wiley, 1968. Pp. xiv, 218. $6.95.

Weiss, James M. A. *Nurses, Patients, and Social Systems.* Columbia: University of Missouri Press, 1968. Pp. 205. $4.50.

West, D. J. *Homo-Sexuality.* Chicago: Aldine, 1968. Pp. 304. $5.95.

Westcott, Malcolm R. *Toward a Contemporary Psychology of Intuition: A Historical, Theoretical, and Empirical Inquiry.* New York: Holt, Rinehart and Winston, 1968. Pp. x, 228. $6.50.

Wicker, Tom, introd. *Report of the National Advisory Commission on Civil Disorders.* New York: Bantom Books, 1968. Pp. xxv, 609. Paperback, $1.25.

Winick, William. *Industry in the Hospital: Mental Rehabilitation through Work.* Springfield, Ill.: Charles C Thomas, 1967. Pp. viii, 84. $5.75.

Wittenberg, Rudolph. *Postadolescence: Theoretical and Clinical Aspects of Psychoanalytic Therapy.* New York: Grune & Stratton, 1968. Pp. vi, 138. $6.50.

Wolman, Benjamin B. *Historical Roots of Contemporary Psychology.* New York: Harper & Row, 1968. Pp. viii, 376.

Wolman, Benjamin B. *The Unconscious Mind: The Meaning of Freudian Psychology.* Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. xi, 244. Cloth, $5.95; paperback, $2.95.

Wright, George N., and Butler, Alfred J. *Rehabilitation Counselor Functions: Annotated References.* Madison: University of Wisconsin Press, 1968. Pp. viii, 451. Paperback.

Wright, George N., and Trotter, Ann Beck. *Rehabilitation Research.* Madison: University of Wisconsin Press, 1968. Pp. xiii, 674. $14.00.

Zubin, Joseph, and Freyhan, Fritz A., eds. *Social Psychiatry.* New York: Grune & Stratton, 1968. Pp. xii, 383. $14.50.

Zubin, Joseph, and Jervis, George A. *Psychopathology of Mental Development.* New York: Grune & Stratton, 1968. Pp. xiii, 658. $22.00.

THE AMERICAN
JOURNAL OF PSYCHOLOGY

Founded in 1887 by G. Stanley Hall

Vol. 82 JUNE 1969 No. 2

ORIGINALITY AND RATE OF RESPONSE IN ASSOCIATION AS A FUNCTION OF ASSOCIATIVE GRADIENT

By Paul V. Olczak and Martin F. Kaplan,
Northern Illinois University

The recent literature has suggested several approaches to the study of originality. One approach defines original thinking as the forming of associative elements into new combinations which meet specified requirements.[1] The more mutually remote the elements, the more original the solution. This definition has been operationalized in the form of a test, the Remote Associates Test, which purports to measure the ability to form new associations into highly original responses.[2] The test orders individuals along a dimension that represents the "steepness" or "flatness" of their associative hierarchies. "Flat" associative hierarchies are those in which the probabilities of evocation of a large number of associates are relatively equal. "Steep" hierarchies are those in which a few associates are extremely probable. Highly original individuals are presumably characterized by flat associative hierarchies, as steep hierarchies represent a lack of originality.

On the other hand, another recent and associative interpretation defines originality as behavior which occurs relatively infrequently but is relevant to given conditions.[3] Thus, an association which is normatively uncommon, *i.e.* of low associative strength, may be considered original. By this logic, it is conceivable that the task of combining disparate elements as required by the Remote Associates

* Received for publication January 13, 1969. Mr. Lester Knapp assisted in instrumentation, and Mr. Daniel Urban helped with the scoring of responses.

[1] S. A. Mednick, The associative basis of the creative process, *Psychol. Rev.*, 69, 1962, 220–232.

[2] Mednick, *loc. cit.*

[3] I. Maltzman, On the training of originality, *Psychol. Rev.*, 67, 1960, 229–242.

Test calls for responses of low associative strength to a given stimulus. The present study was an exploration of several aspects of this possibility.

Garskof, Shapiro, and Brandstadter have found that continued association yields associative-hierarchy data similar to those obtained in conventional associative-frequency data.[4] Mednick, Mednick, and Jung have found that individuals with high scores on the Remote Associates Test showed a greater quantity of continued associations than low scorers.[5] Their study also showed that frequently occurring words elicited more responses than words of low frequency, flat-hierarchy words elicited more responses than steep-hierarchy words (see below), and nouns more responses than adjectives. This suggested that a continued-association task might be a useful idiographic approach to word-association data in the present study. One easily obtainable measure of associative strength would be associative reaction time.[6] It would be expected that associations low in the hierarchy would require longer latencies due to their relative inaccessibility. Further, flat associative hierarchies should be characterized by longer latencies due to inherent response competition. The first hypothesis, therefore, was that if high scorers on the Remote Associates Test possess flatter hierarchies, they should evidence initially longer reaction times in continued association than low scorers, who presumably possess fewer but more dominant associations. As continued association progresses and low scorers exhaust their more dominant responses, response competition should increase and latencies for these later trials should approach those of high scorers.

As it is possible to characterize individuals as having flat hierarchies or steep hierarchies, it is possible to so characterize stimuli. Stimuli which elicit one dominant association and many associations of low response strength may themselves be considered to have steep associative hierarchies, while those which do not elicit any given dominant response may be considered to have flat associative hierarchies.[7] The second hypothesis, therefore, was that the initially

[4] B. E. Garskof, Relation between single word association and continued association response hierarchies, *Psychol. Rep.*, 16, 1965, 307–309; B. E. Garskof, E. B. Shapiro, and J. Brandstadter, Order of emission in continued association as a predictor of individual free recall, *Psychon. Sci.*, 1, 1967, 209–210.

[5] M. T. Mednick, S. A. Mednick, and C. C. Jung, Continual association as a function of level of creativity and type of verbal stimulus, *J. abnorm. soc. Psychol.*, 69. 1964, 511–515.

[6] R. Wallenhorst, Some relations between reaction time and choice of response in word association, *Psychol. Rep.*, 17, 1965, 619–626.

[7] Mednick, Mednick, and Jung, *loc. cit.*

greater response competition reflected in flat-hierarchy words should result in initially greater reaction time than steep-hierarchy words, differences between the words decreasing as the few dominant responses to the steep-hierarchy words are exhausted in continued association. In an attempt at generality, both hypotheses were tested using both word-association and Thematic Apperception Test (TAT) stimuli, the latter representing a more complex associative task.

Method

Subjects. The *S*s were chosen from 300 women enrolled in an introductory psychology course, all of whom had been given the Remote Associates Test. On the basis of their scores being one *SD* above or below the sample mean, 25 *S*s were chosen as High Scorers (score $>$ 20) and 25 *S*s as Low Scorers (score $<$ 11).

Materials. Word-association stimuli were selected from the Palermo and Jenkins norms.[8] There were 20 words selected; all were nouns, 10 of them steep in associative strength (steep-hierarchy, or steep-slope, words) and 10 flat in associative strength (flat-hierarchy, or flat-slope, words) following Mednick, Mednick, and Jung's definitions of word slope.[9] The two sets of words were equated for Thorndike-Lorge frequency.[10]

Associative slope of TAT stimuli was determined by reference to Eron's norms.[11] Cards of steep associative slope (steep-hierarchy, or steep-slope, cards) were defined as those in which the most dominant theme appeared with greater than 35% frequency in the responses of the normative group and with at least 20% greater frequency than the next most popular theme. Cards of flat associative slope (flat-hierarchy, or flat-slope, cards) were defined as those for which no theme appeared with greater than 34% frequency and in which the two most dominant themes were within 20% frequency of one another. On this basis, Cards 3-GF and 12-F were considered flat-slope cards and Cards 13-MF and 19 were considered steep-slope cards.

*Word association. S*s were asked to respond in continued association to each of the 20 stimulus words, the order of which was individually randomized, with six associates. *S*s were instructed not to repeat responses on successive associations to the same word. To minimize chaining, the word was presented each time. Stimuli were presented, via Koss Pro-4A earphones, by *E* from an adjoining room. *E*'s oral presentation of the word activated a timer (Transistor Specialties, Inc., Model 361-R), which was stopped by *S*'s response. Responses were taped and monitored by a SONY Model TC-200 tape recorder.

[8] D. S. Palermo and J. J. Jenkins, *Word Association Norms: Grade School through College*, 1964.

[9] Flat-hierarchy words were *baby, cheese, people, whistle, butterfly, child, street, cars, memory, doctor;* steep-hierarchy words were *black, table, butter, blossom, boy, lamp, man, scissors, woman, bed.* See Mednick, Mednick, and Jung, *loc. cit.*

[10] E. L. Thorndike and I. Lorge, *The Teacher's Word Book of 30,000 Words*, 1944.

[11] L. D. Eron, Responses of women to the TAT, *J. consult. Psychol.*, 17, 1953, 269–282.

Thermatic Apperception Test. Following the word-association task, E entered the room with S and gave these instructions:

I will present you with a picture to which you will tell six different stories. The stories should be only a few sentences long, telling who the characters are and what they are doing. Do not just describe the picture, but try to tell a story. When you receive a picture, press the key down. Keep the key down while you are thinking of a story. Release the key when you begin telling the story. When you end that story, again press the key down as you think of your next story. In general, when you are thinking of a story, the key is down; when you are telling a story, the key is up. Try to give a different story each time. Do not worry about counting the number of stories you have given. You will be notified when you have given six stories.

A Hunter Klock Kounter (Model 120-A, Series D) was activated by depression of the telegraph key, thus obtaining elapsed time. Themes were also recorded on tape; the tape recorder and timing mechanism were located in an adjoining room.

Results

Fig. 1 presents mean reaction times for the two sets of words for each of the six trials for the two groups of Ss (High Scorers and Low Scorers). The data were analyzed by means of a $2 \times 2 \times 6$ repeated-measures analysis of variance design, with Ss nested under levels of originality as measured by the Remote Associates Test.[12] The trials effect was significant ($F = 49.36$, df 5/240, $p < .001$), as was the Slope of Words $\times$ Trials interaction ($F = 4.65$, df 5/240, $p < .01$). Reaction times increased as trials progressed, and flat-slope words tended to evoke longer reaction times than steep-slope words on Trial 1, the differences between the words becoming negligible on succeeding trials. Trends approaching significance were noted for flat-slope words to evoke longer reaction times than steep-slope words ($F = 3.12$, df 1/48, $p < .10$) and for High Scorers to respond faster than Low Scorers ($F = 3.38$, df 1/48, $p < .10$).

Fig. 2 shows the mean response latencies for the two sets of TAT cards for each of the six trials for the two groups of Ss. The following effects were significant: trials ($F = 11.85$, df 5/230, $p < .01$) and slope of cards ($F = 7.74$, df 1/46, $p < .01$). The F ratio for differences between High and Low Scorers was found to be less than one. Again, reaction times generally increased as trials progressed. The flat-slope cards obtained shorter latencies than the steep-slope cards.

Responses to the word-association task were scored for originality by assigning each response the frequency of occurrence for that associate in the Palermo and Jenkins norms, so that the lower the score, the more novel the response (see Fig. 3).[13] The following ef-

[12] B. J. Winer, *Statistical Principles in Experimental Design*, 1962, 319–337.
[13] Palermo and Jenkins, *op. cit.*

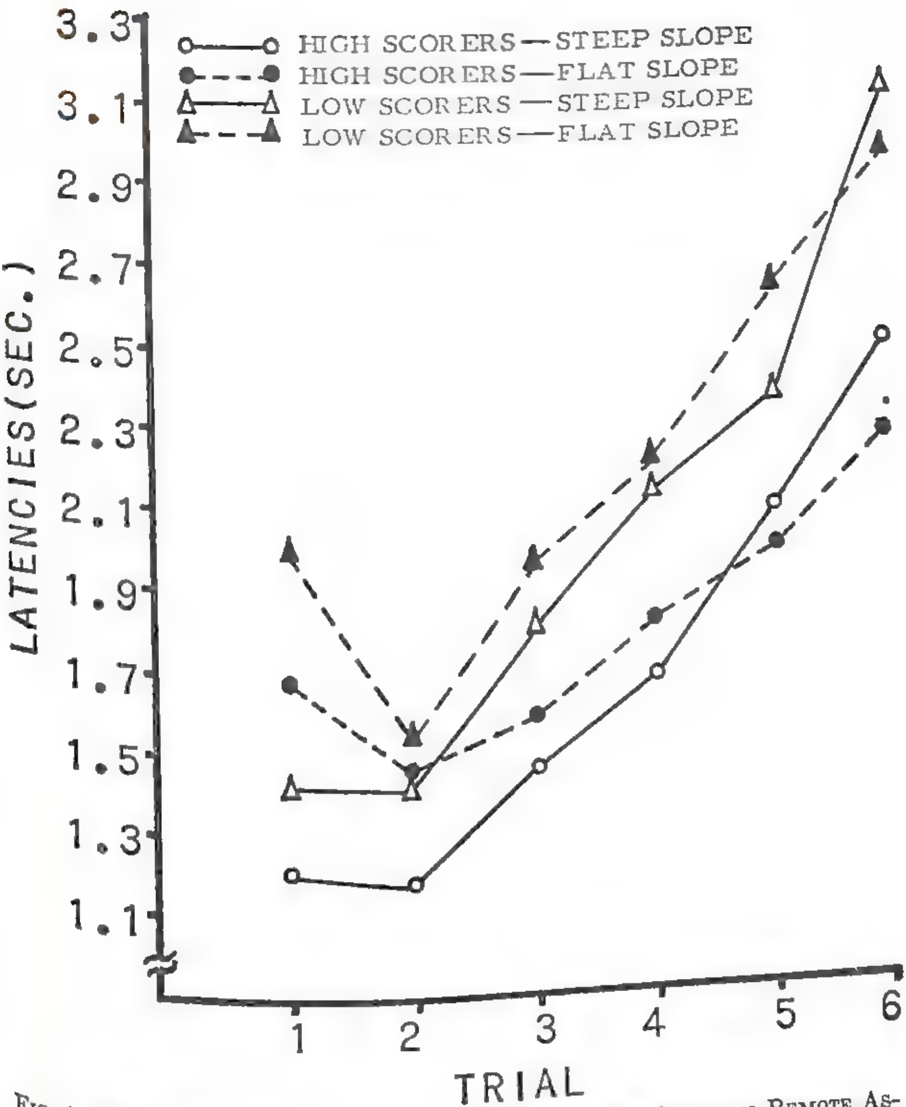


FIG. 1. WORD-ASSOCIATION LATENCIES AS A FUNCTION OF SCORE ON REMOTE ASSOCIATES TEST AND ASSOCIATIVE SLOPE OF WORDS

fects were significant: trials ($F = 4.95$, df 5/240, $p < .01$) and slope of words ($F = 17.10$, df 1/48, $p < .01$). In general, responses became more original as trials progressed; this effect was largely attributable to the fact that the steep-slope words initially elicited very common responses, the differences between the two sets diminishing on subsequent trials as in Fig. 1.

Responses to the TAT were similarly scored for originality by assigning each of the six themes for each card the frequency of occurrence of that theme in Eron's norms, so that lower scores in-

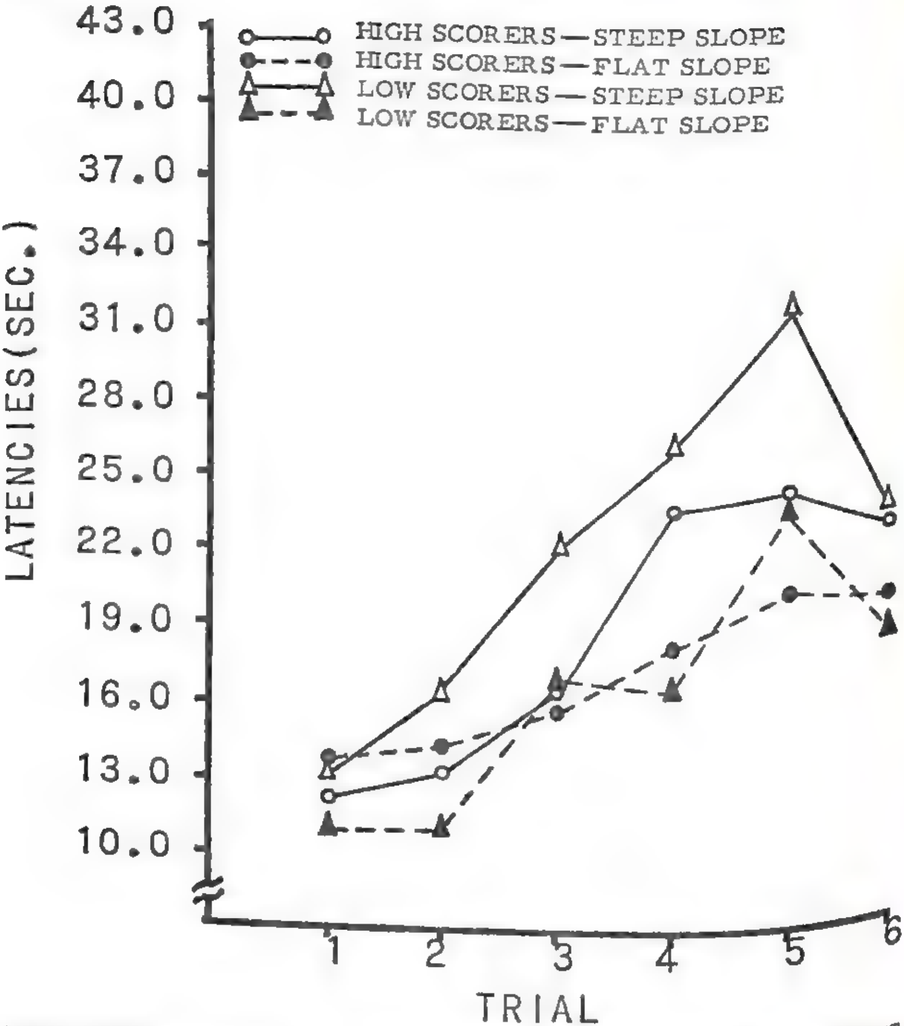


FIG. 2. TAT LATENCIES AS A FUNCTION OF SCORE ON REMOTE ASSOCIATES TEST AND ASSOCIATIVE SLOPE OF CARDS

dicate greater infrequency or novelty of response.[14] Themes were scored by reference to Eron's checklist, and reliability of scoring was estimated by percentage of agreement between two judges. Of 196 themes sampled, the two judges agreed with the scoring of the themes on 162, or 83%, of them. This reliability estimate compares favorably with estimates derived from other studies using identical novelty scores.[15] Fig. 4 presents the mean novelty scores for the two

[14] Eron, *loc. cit.*

[15] M. F. Kaplan and L. D. Eron, Test sophistication and faking in the TAT situation, *J. proj. Tech.*, 29, 1965, 498–503; M. Kaplan, S. Simon, and R. Ditrichs,

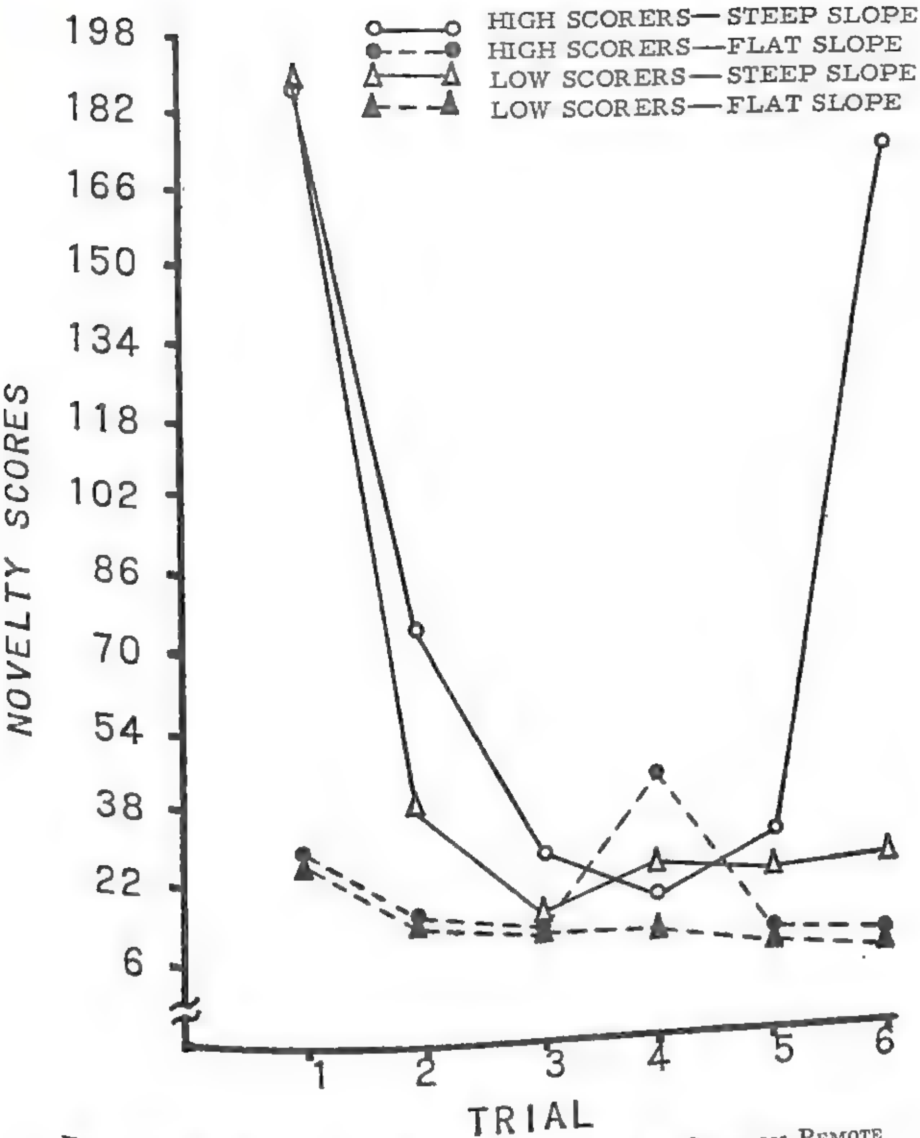


FIG. 3. WORD-ASSOCIATION NOVELTY AS A FUNCTION OF SCORE ON REMOTE ASSOCIATES TEST AND ASSOCIATIVE SLOPE OF WORDS

sets of cards for each of the six trials for the two groups of *S*s. The following effects were significant: trials ($F = 13.65$, df 5/230, $p < .01$), slope of cards ($F = 317.11$, df 1/46, $p < .001$), and Slope of Cards × Trials interaction ($F = 10.33$, df 5/230, $p < .01$). Again, responses became more original as trials progressed and

Repression-sensitization as a variable in the observational learning of TAT themes, unpublished study, 1968.

there were no significant differences between High and Low Scorers. The steep-slope cards were responded to with more common themes throughout than the flat-slope cards, with differences diminishing over trials.

Discussion

The finding of general interest is the fact that both latency and novelty, or uncommonality, of response increased as trials progressed. This is consistent with the concept that potential associations to a stimulus form a hierarchy of associational strength, un-

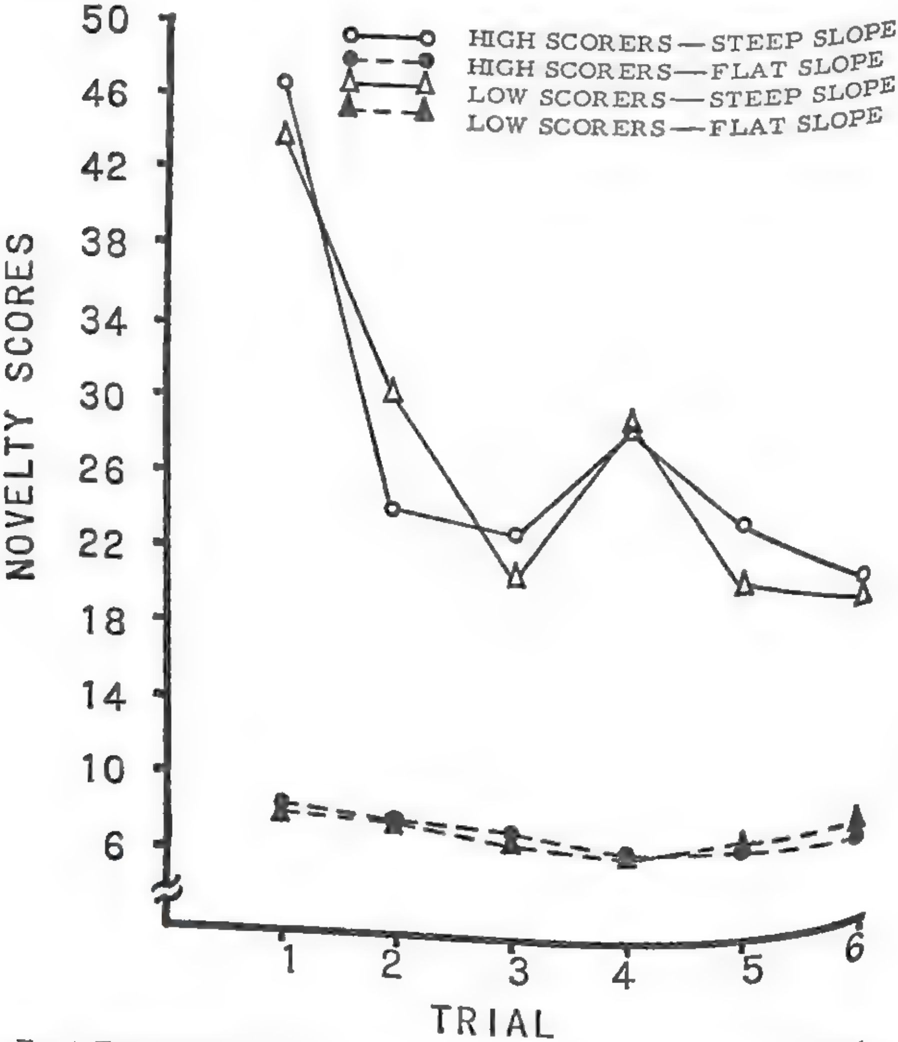


Fig. 4. TAT Novelty as a Function of Score on Remote Associates Test and Associative Slope of Cards

common associations being responses with lower probabilities of evocation. Apparently, as continued-association trials progress, and more novel responses are emitted, reaction time increases. This finding supports Wallenhorst's contention that reaction time is a suitable means of measuring response probability and is also consistent with the finding by Garskof, Shapiro, and Brandstadter that order of emission in continued association is an index of slope of the associative hierarchy.[16] The results of the present study suggest that reaction time in continued association can also be used to assess slope of the associative hierarchy in a more complex associative task, *i.e.* the TAT. It was predicted that stimuli with normatively different slopes would elicit different patterns of reaction time in continued association. This prediction was supported when words were used as stimuli. Flat-slope words, due to initially high response competition, obtained longer latencies than did words with steep slopes, *i.e.* with one dominant response. As trials progressed, and the dominant response was eliminated, competition in steep-slope words increased and reaction-time differences between the two sets of words decreased (see Fig. 1). The results were not as clear-cut when TAT cards were used as stimuli. Steep-slope cards generally elicited longer reaction times than flat-slope cards, but as Fig. 2 shows, this difference occurred only on Trials 2–6, the reaction times being similar on Trial 1. Evidently the steep-slope cards obtained more response competition than did the flat-slope cards, except on Trial 1, where a dominant response was present. That reaction time with steep-slope TAT cards was not initially faster, as it was for steep-slope words, may be due to the fact that the TAT stimuli are more complex and ambiguous, and even relatively dominant responses are not free from a good deal of response competition.

Mednick suggests that creativity, and by implication originality, is associated with flat associative hierarchies.[17] In general, flat-slope stimuli elicited more original responses than did steep-slope stimuli; this was true of both the word-association and TAT tasks. The difference between flat-slope and steep-slope stimuli was greatest in the earlier trials, diminishing as trials progressed. That is, as the more common, dominant responses in the steep-slope hierarchies were depleted, responses became more normatively uncommon. For the flat-slope stimuli, on the other hand, responses were initially un-

[16] Wallenhorst, *loc. cit.*; Garskof, Shapiro, and Brandstadter, *loc. cit.*
[17] Mednick, *loc. cit.*

common and remained so as trials progressed. Highly competitional stimuli, therefore, were more likely to elicit original responses, and more likely do so on early trials, where the common, dominant responses to less competitional stimuli were not yet depleted. It should be noted that continued association resulted in little or no increase in originality with either kind of flat-slope stimuli, either cards or words. Therefore, the continued-association method of training for originality employed by Maltzman[18] is likely to be effective only with steep-slope stimuli.

While slope of associative hierarchy for the stimuli was found to be a significant factor in both reaction time and novelty of response, associative slope for subjects was not. Except for the fact that the High Scorers tended to respond faster than the Low Scorers in word association, no differences between the groups were found. Thus, to the extent that the Remote Associates Test is a measure of associative slope, the presence of typically "steep" or "flat" slopes in people is not analogous to differential slope of associations for stimuli. These findings suggest that something other than slope of hierarchy is being measured by the Remote Associates Test. Perhaps the most salient feature being tapped is extent of hierarchy; Mednick, Mednick, and Jung, for instance, have shown that High Scorers possess more continued associations.[19] This could account for observed differences in reaction time, but these differences were limited to word association. Another plausible explanation for the negative results encountered with the Remote Associates Test is that in trying to assess a presumably general and pervasive personality trait such as slope of associative hierarchy, a more diverse sample of behavior than the test affords is required. A third possibility is, perhaps, that people cannot be characterized as having typically steep or flat associative hierarchies in a chronic sense, but that the slope of a given individual's hierarchy depends on the particular stimulus context; *i.e.* that 'slopeness' is not an idiosyncratic personality trait. Future study should determine whether individuals tend to evidence characteristic slopes of associative strengths to varied stimuli.

SUMMARY

Reaction time and novelty of response to both words and TAT cards as stimuli were studied as a function of normative slope of

[18] Maltzman, *loc. cit.*
[19] Mednick, Mednick, and Jung, *loc. cit.*

the associative strength of generally elicited responses. *S*s presumably high or low in steepness of associative slope as measured by the Remote Associates Test responded in continued associations to words and TAT cards normatively possessing either "flat" or "steep" associative slopes. In general, reaction time and novelty of response increased over trials.

More novelty of response was elicited by both types of flat-slope stimuli, and the difference between flat-slope and steep-slope stimuli was greatest on the earlier trials, where the steep-slope stimuli elicited relatively common responses. For reaction time, a similar observation was made in the word-association task: flat-slope words initially resulted in longer reaction times, but the difference between flat-slope and steep-slope words decreased as trials progressed. These results were seen as consistent with the notion that reaction time can serve as a measure of relative response competition and associative slope. Further, the ordering of stimuli into flat and steep associative slopes seemed justified by both reaction time and novelty data; and the equating of originality with associative slope received some support.

Scores on the Remote Associates Test were not a significant variable, and three possible conclusions were advanced. (1) The test may not measure a quality in people which is analogous to associative slope of stimuli. (2) Although a generalized, chronic personality trait such as slope of associative hierarchy may exist, the test may not be a broad enough sample of representative behavior. (3) Slope of associative hierarchy may not be a generalized personality trait, *i.e.* may not remain constant in a given individual over a variety of stimulus contexts but, rather, may vary with a given stimulus-person interaction.

COGNITIVE PROCESSES IN DIFFERENTIAL GSR CONDITIONING: EFFECTS OF A MASKING TASK

By Marcus J. Fuhrer and Paul E. Baer,
Baylor University College of Medicine

A series of studies has demonstrated that acquisition of differentially conditioned GSRs by intellectually capable human *Ss* is associated with verbalized recognition of the CS-UCS contingencies.[1] In these studies, conditional GSR differentiation was acquired only by *Ss* who could accurately verbalize the stimulus relations during either intertrial verbal reports or postconditioning interviews. A similar relationship between knowledge of stimulus relations and acquisition of classically conditioned responses has also been demonstrated by studies involving informative preparatory instructions: Verbal description of the stimulus contingencies prior to conditioning has been shown to facilitate acquisition of conditional GSRs,[2] heart rate,[3] and eye blinks.[4] It follows that acquisition of

* Received for publication July 26, 1968. The investigation reported was supported in part by Public Health Service Grants MH 11939-01 and MH 12908-01; the Social and Rehabilitation Service, Department of Health, Education and Welfare, Research and Training Center RT-4; and the Texas Research Institute of Mental Sciences. The Baylor University College of Medicine Computer Science Program and the Common Research Computer Facility of the Texas Medical Center assisted in data analyses. The first author (Marcus J. Fuhrer) is also at the Texas Institute for Rehabilitation and Research.

[1] M. J. Fuhrer and P. E. Baer, Differential classical conditioning: Verbalization of stimulus contingencies, *Science*, 150, 1965, 1479–1481; M. J. Fuhrer and P. E. Baer, Effects of intertrial reports on cognitive and GSR differentiation of conditional stimuli, *Proc. 74th Annu. Conv. Amer. Psychol. Ass.*, 1966, 53–54; P. E. Baer and M. J. Fuhrer, Cognitive processes during differential trace and delayed conditioning of the GSR, *J. exp. Psychol.*, 78, 1968, 81–88.

[2] S. W. Cook and R. E. Harris, The verbal conditioning of the galvanic skin reflex, *J. exp. Psychol.*, 21, 1937, 202–210; J. D. Block, Awareness of stimulus relationships and physiological generality of response in autonomic discrimination, in J. Wortes (ed.), *Recent Advances in Biological Psychiatry*, 1962, 43–53.

[3] B. B. Chatterjee and C. W. Eriksen, Cognitive factors in heart rate conditioning, *J. exp. Psychol.*, 64, 1962, 272–279; M. J. Fuhrer, Differential verbal conditioning of heart rate with minimization of changes in respiratory rate, *J. comp. physiol. Psychol*, 58, 1964, 283–289.

[4] E. R. Hilgard, A. A. Campbell, and W. N. Sears, Conditioned discrimination: The effect of knowledge of stimulus relationships, this Journal, 51,

CRs should be impeded if the conditioning procedure is embedded in a masking task designed to interfere with recognition of stimulus relations. There have been relatively few tests of this clear-cut expectation since available studies using a masking task have been chiefly concerned with extinction processes in eye-blink conditioning,[5] but some corroborative evidence has been reported by Ross, Wilcox, and Mayer, who found that differential eye-blink conditioning is markedly impaired when a time-estimation task is used as a masking procedure.[6] With respect to GSR conditioning, Dawson and Grings have reported an absence of conditional GSR differentiation for *S*s who took a paper-and-pencil test during conditioning and who were subsequently unable to describe the stimulus relations accurately.[7] Dawson and Grings note, however, that this masking task may have interfered with perception of the individual CSs rather than with perception of the CS-UCS relations.

The primary purpose of the present study was to determine how differential GSR conditioning is affected by a masking task which ensures perception of the CSs. The task, a probability-learning procedure, was similar to one used by Spence.[8] A group which had the conditioning procedure embedded in the masking task (Group M) was compared with a group which underwent a conventional conditioning procedure (Group C). In both groups, a detailed postconditioning interview was used to ascertain which *S*s could accurately describe the conditional stimulus relations. On the basis of previous findings in this series of studies, it was expected that accurately verbalizing *S*s in Group C would differentially condition but that inaccurately verbalizing *S*s would not. Because perception of the CSs was assured in Group M, it was anticipated that some *S*s would be able to verbalize the stimulus relations accurately. It could then be determined whether similar relationships between accuracy of verbalization and GSR conditioning would be found for the M and C groups. Because a CS-UCS interval of 8 sec. was used,

---

1938, 598–606; W. R. McAllister and D. E. McAllister, Effect of knowledge of conditioning upon eyelid conditioning, *J. exp. Psychol.*, 55, 1958, 579–583.

[5] K. W. Spence, Cognitive and drive factors in the extinction of the conditioned eye blink in human subjects, *Psychol. Rev.*, 73, 1966, 445–458.

[6] L. E. Ross, Susan M. Wilcox, and Melanie J. Mayer, Simple eyelid conditioning masking task and its effect on differential conditioning, *Psychon. Sci.*, 9, 1967, 333–334.

[7] M. E. Dawson and W. W. Grings, Comparison of classical conditioning and relational learning, *J. exp. Psychol.*, 76, 1968, 227–231.

[8] K. W. Spence, Cognitive factors in the extinction of the conditioned eyelid response in humans, *Science*, 140, 1963, 1224–1225.

evidence for differential conditioning was assessed separately in three response intervals defined by the latency of the GSR.[9]

METHOD

Subjects, stimuli, and apparatus. College-age men recruited from the University of Houston served as paid volunteers. There were 43 *Ss* in Group M and 24 *Ss* in Group C. Exosomatic skin resistance and GSRs were recorded monopolarly using a constant voltage (.4 v.) system[10] in conjunction with an Offner Type-R Dynograph. Basal resistance was measured in ohms and the GSR in micromhos. The CSs were 8-sec. tones of 700 and 3400 cps. The tones had been psychophysically matched for loudness by another group of *Ss* and were clearly audible. The UCS was a .25-sec., 100 pulse/sec shock supplied by an E & M stimulator and isolation unit applied to the left forefinger. The UCS onset was coincident with offset of the tone, the duration of all stimuli being controlled by Hunter timers. A panel located 48 in. in front of *S* contained two white lights, horizontally separated by 12 in. The lights were at *S*'s eye level and could be separately actuated by *E* for a 1-sec. period.

Procedure. S was seated in a sound-attenuated, semidarkened room adjacent to the instrumentation room. Punched masking tape was applied to the palmar surface of the second phalange of the second finger of the right hand to demarcate the active site. An Ag-AgCl electrode was coated with a gel which is isotonic with sweat and taped over the active site. The 10 × 8 cm. reference electrode was coated with gel and taped to the ventral surface of the forearm. The UCS intensity was then adjusted to *S*'s tolerance level.

Ss in Group M were told that the study was concerned with problem-solving in the face of distraction. The task was described as one which required guessing whether a light on the left or right would be illuminated. *S* was told that a tone in the earphones would be his signal to formulate his guess and his confidence in the accuracy of his guess on an eight-point scale, and, also, that several seconds after offset of the tone, *E* would say "guess" over the intercom as a signal for *S* to state his guess and confidence level. It was suggested to *S* that he could improve the accuracy of his guesses, even though a distracting shock would be occasionally administered. *Ss* in Group C were simply told that they would occasionally receive a shock during the experiment.

During the 10 adaptation trials, equal numbers of the two tones were presented in an unsystematic order. During acquisition the frequency of the tone which was paired with shock (CS+) was counterbalanced within each group. The 33 acquisition trials consisted of 17 presentations of the CS+ tone and 16 presentations of the unpaired tone (CS−). Test trials were administered by omitting the UCS during Presentations 3, 7, 11, and 15 of the CS+. The CS+ and CS− trials were in an unsystematic order except for the restrictions that the same tone not occur more than twice successively and that each test trial be bracketed by CS− tones. The intertrial interval varied unsystematically

[9] M. A. Stewart, J. A. Stern, G. Winokur, and S. Fredman, An analysis of GSR conditioning, *Psychol. Rev.*, 68, 1961, 60–67; W. F. Prokasy and H. C. Ebel, Three components of the classically conditioned GSR in human subjects, *J. exp. Psychol.*, 73, 1967, 247–256.

[10] J. D. Montagu and E. M. Coles, Mechanism and measurement of the galvanic skin response, *Psychol. Bull.*, 65, 1966, 261–278.

between 20 and 40 sec. for both groups. A 7:3 ratio was used for presentation of left and right lights for *Ss* in Group M. After the last acquisition trial, all *Ss* were asked these increasingly specific questions:

1. I want you now to tell me about your experience during the experiment. Review in detail your observations beginning at the first of the experiment and continuing through to its end. In a sense, this is a test of your ability to communicate fully what you observed, thought, and did during the experiment. Take your time and try to be as complete in your description as possible.

2. It will help you to recall more details of your thoughts and observations during the experiment by concentrating on one topic at a time. So, to start, please trace your thoughts and observations about the shocks you received from the beginning to the end of the experiment. Some of what you say may repeat what you have already told me, but that is all right.

3. Now, I would like you to tell me about your experience in the experiment in terms of the tones you heard. So that you can give me a detailed account, again try to think back to the first tone you heard and trace your observations and thoughts about the tones from the beginning to the end of the experiment.

4. We know that *Ss* attempt to organize what is going on during the experiment in some manner. However, each person does this a little differently. Please tell me in detail how you organized and attempted to make sense about what was happening in the experiment.

5. As a result of your experience in the experiment, what conclusions did you come to concerning the experiment? That is, as a person undergoing the experiment, how did you try to explain the experiment to yourself?

6. You noticed that there were repeated shocks during the experiment. What sense did you make out of the fact that you got a shock at a particular time?

7. Were you ever able to tell when you were going to be shocked? If so, how?

8. Did you notice any connection between the tones and the shock?

9. Did you ever notice yourself anticipating the shock when you heard a tone?

10. Did you make any sense out of the fact that sometimes there was a tone which was not followed by shock?

11. How many types of tones did you hear?

12. [If more than one.] Please describe the types of tones.

13. Did hearing a particular type of tone ever lead you to predict whether you would be shocked or not? [If yes.] Which type of tone? Were you clearly aware of this during the experiment? Estimate when in the experiment you thought this. Did you modify or possibly forget this idea after you first got it?

14. [An *S* who claimed not to have heard the tones or differentiated their pitches was tested at this point by presenting both tones and obtaining the *S*'s judgments.]

Transcripts of these postconditioning interviews for both the M and C groups were then independently content-analyzed by four raters for the accuracy of *Ss*' conceptualizations of the conditional stimulus contingencies. The raters dealt with these questions:

1. How many types of tones did the *S* report hearing?
2. Did the *S* describe the tones as differing in pitch?
3. Did the *S* report associating tones and shocks?
4. Did the *S* report associating shocks with just one type of tone?
5. Did the *S* report that one type of tone was *not* associated with shock?
6. Did the *S* report having rejected the contingency after having recognized it?
7. Did the *S* report attempting to predict shock on a basis other than the CS-UCS contingency: (*a*) sequence of shock and nonshock; (*b*) stimuli other than the CS; (*c*) responses other than CR or UCR; (*d*) other?

Note: Rate "SC" if the S's confidence in his observation is qualified. Rate "RC" if your confidence as a rater is qualified.

The raters had no knowledge of the GSR conditioning obtained for each S. Ratings of Ss' responses to Question 13 were not used to classify Ss. Assignment of S to an accurately verbalizing group required either three unqualified ratings of accuracy or four ratings of accuracy if any rater considered his rating qualified. Unanimity was achieved among the four raters for 24 of the 28 Ss who were classified as accurate verbalizers in the two groups.

The 8-sec. interval beginning 1 sec. after tone onset was divided into two subintervals. A GSR originating 1–5 sec. after tone onset was scored as a first-interval response. A GSR originating 5–9 sec. after tone onset was scored as a second-interval response. Thus, 1 sec. was alloted as the minimum latency of a GSR to the UCS. For trials in which the UCS was not administered, a third-interval response was scored for a GSR originating 9–13 sec. following tone onset. Each GSR was measured from its onset to its peak, and the magnitude was expressed as the square root of the conductance change.

Results

In Group M, 13 Ss were classified as accurate and 30 as inaccurate; in Group C, 15 Ss were classified as accurate and 9 as inaccurate, $\chi^2 = 5.3$, $p < .05$, two-tailed.

First-Interval GSRs. The upper left panel of Fig. 1 shows the mean first-interval GSR magnitudes to CS+ and CS− during the

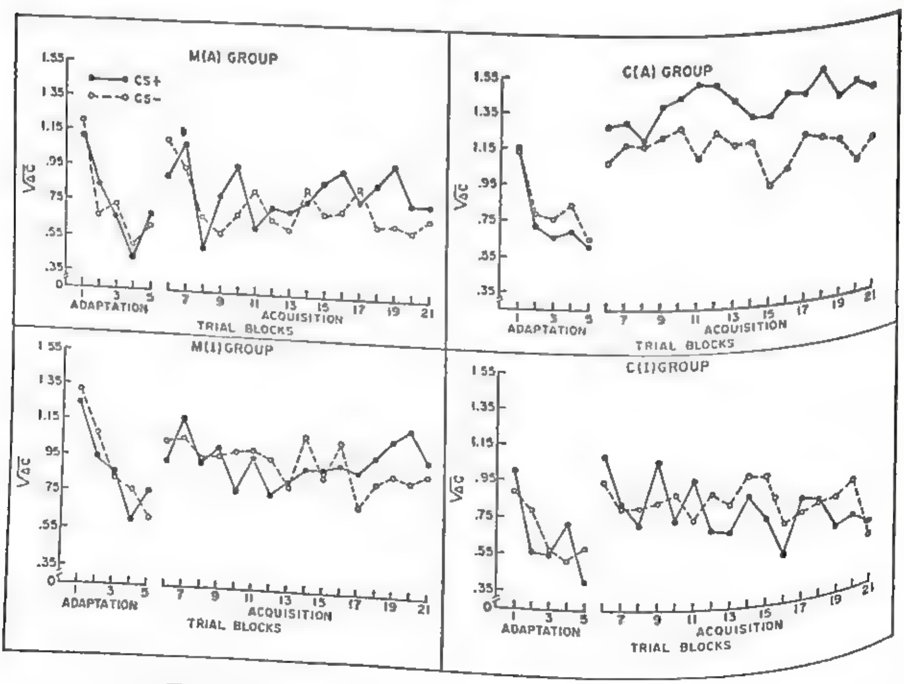


Fig. 1. Mean First-Interval GSR Magnitudes

blocks of adaptation and acquisition trials for the 13 *Ss* in Group M who accurately verbalized the contingencies. These *Ss* were designated Group M(A). The lower left panel contains comparable data for the 30 inaccurate verbalizers, designated Group M(I). Trial blocks were comprised of the most nearly adjacent pair of CS+ and CS− trials. The first CS+ trial during acquisition was omitted since GSRs elicited by this stimulus could not reflect sensitization induced by UCS presentations. Three-factor analyses of variance—the M(A) vs. the M(I) group, CS+ vs. CS−, and trial blocks—were calculated separately for blocks of adaptation and acquisition trials. The M(A) and M(I) groups did not differ in GSR magnitudes to CS+ and CS− combined during adaptation, $F(1, 41) = 1.1$, and the combined groups showed a highly significant degree of habituation, $F(4, 164) = 33.7, p < .001$. Analysis of blocks of acquisition trials failed to demonstrate a main effect for CS+ vs. CS−, $F(1, 41) = 1.5$. However, a significant interaction between CS+ vs. CS− and trial blocks was obtained, $F(15, 615) = 1.8, p < .05$, suggesting that conditional GSR differentiation occurred for the M(A) and M(I) groups combined as a function of trial blocks. As indicated in Fig. 1, neither group began to differentiate consistently until the relatively late acquisition trials. Analyses of difference scores reflecting differentiation of the combined groups over Trial Blocks 17–21 revealed significant conditioning, $t(42) = 4.0, p < .001$. The two groups did not differ in overall first-interval GSR differentiation, $F(1, 41) = 1.2$, nor did the differentiation of the groups differ as a function of trial blocks, $F < 1$. Indeed, the M(A) and M(I) groups both differentiated significantly during Trial Blocks 17–21, $t(12) = 2.3, p < .05$; $t(29) = 3.2, p < .01$, respectively. To test the possibility that the basis for assigning *Ss* to the M(I) group might account in part for this group's GSR differentiation during Trial Blocks 17–21, the 30 *Ss* were partitioned into two subgroups. One subgroup consisted of 13 *Ss* judged by all four raters as having failed to verbalize the stimulus relations accurately in response to all interview questions. The other subgroup was comprised of 17 *Ss* who obtained some ratings of accuracy but did not meet the strict criteria required for classification in Group M(A). There was no tendency for the GSR differentiation of these subgroups to differ, $t(28) = .59$, and each subgroup separately showed significant differentiation, $t(12) = 2.34, p < .05$; $t(16) = 2.67, p < .05$.

The upper right panel of Fig. 1 shows the mean first-interval GSR

magnitudes for the 15 *Ss* in Group C who accurately verbalized the contingencies. These *Ss* are designated Group C(A). Comparable data for the 9 inaccurate verbalizers, designated as Group C(I), are shown in the lower right panel. The reactivity of the groups to CS+ and CS− combined during adaptation did not differ, $F(1, 22)$, $p < 1$, and the combined groups significantly habituated to CS+ and CS− over the blocks of adaptation trials, $F(4, 88) = 12.6$, $p < .001$. During acquisition, significant GSR conditioning was obtained for C(A) and C(I) groups combined, $F(1, 22) = 11.8$, $p < .01$. As shown in the right-hand panels of Fig. 1, the groups differed markedly in conditional GSR differentiation, $F(1, 22) = 11.9$, $p < .01$, with conditional differentiation being limited to the C(A) group. There was also a nonsignificant tendency for the Group C(A) to show larger GSR magnitudes to CS+ and CS− combined than Group C(I), $F(1, 22) = 3.7$, $p < .10$.

The first-interval GSRs of Groups M and C were also compared during blocks of adaptation and acquisition trials, irrespective of *Ss'* accuracy of verbalization. Statistically significant differences between the groups were limited to the blocks of acquisition trials; a significant interaction was obtained between groups and CS+ vs. CS−, $F(1, 65) = 4.2$, $p < .05$. As suggested in Fig. 1, this finding reflected the greater differential conditioning of Group C.

Second-Interval GSRs. The distribution of second-interval magnitudes dictated nonparametric analyses. Two-tailed tests were used throughout. Separate analyses were conducted for the blocks of adaptation trials, the first half of acquisition (Trial Blocks 6–13), and the second half of acquisition (Trial Blocks 14–21). For each half, the mean difference between second-interval GSR magnitudes to CS+ and CS− was calculated separately for each *S*. The median difference scores for the M(A) and M(I) groups, as well as for the two combined (Group M), are shown in Table I. The Wilcoxon test yielded no evidence of statistically significant conditional differentiation during the first half of acquisition for Group M(A), $T(13) = 45$, for Group M(I), $T(30) = 195$, $z = .77$, or for the two combined, $T(43) = 401$, $z = .87$. There was, however, statistically significant differentiation during the last half of acquisition for Group M(A), $T(13) = 7$, $p < .01$, but not for Group M(I), $T(30) = 225$, $z = 0.18$. Use of the Mann-Whitney test indicated that the conditional differentiation of Group M(A) was significantly greater than for Group M(I), $U(13, 30) = 79.5$, $z = 3.0$, $p < .01$.

TABLE I

Median Second-Interval Difference Scores During Adaptation and Acquisition Trial Blocks

Group	Accuracy of verbalization	N	Trial blocks: Adaptation 1–5	Acquisition 6–13	Acquisition 14–21
Masked (M)	Accurate M(A)	13	−.03	.02	.21**
	Inaccurate M(I)	30	.06	−.03	.00
	M(A) and M(I) combined	43	.03	−.04	.04
Comparison (C)	Accurate C(A)	15	.02	.08	.23**
	Inaccurate C(I)	9	−.08	.00	.09
	C(A) and C(I) combined	24	−.01	.06	.16**

** $p < .01$.

Group M as a whole showed only a tendency toward differentiation, $T(43) = 332, z = 1.70, p < .10$.

The median second-interval difference scores for the C(A) and C(I) groups, as well as the two combined (Group C), are also shown in Table I. There was no statistically significant differentiation during the first half of acquisition for Group C(A), $T(15) = 26$, Group C(I), $T(9) = 16$, or for Group C, $T(24) = 86$. For the second half of acquisition, however, there was highly significant evidence of conditional GSR differentiation for Group C(A), $T(15) = 0, p < .01$, for Group C, $T(24) = 14, p < .01$, but not for Group C(I), $T(9) = 10$. In addition, the differentiation of Group C(A) was significantly greater than for Group C(I), $U(15, 9) = 28, p < .02$. The M and C groups were also compared for the two halves of acquisition. Group C differentiated significantly more during the first half, $U(43, 24) = 340, z = 1.96, p < .05$, and during the second half, $U(43, 24) = 314, z = 2.64, p < .01$.

Third-Interval GSRs. Test trials in Trial Blocks 8 and 12 were combined to represent the first half of acquisition, and test trials in Trial Blocks 16 and 20 were combined to represent the second half. Third-interval GSRs in Trial Blocks 1 and 4 were combined to represent the adaptation phase. For each pair of trial blocks, the mean difference score between CS+ and CS− was calculated separately for each *S*. The median difference scores for Group M(A) are shown in Table II. These difference scores were not statistically significant for either the first half of acquisition, $T(13) = 32$, or for the second half, $T(13) = 38$. Nor were the median difference scores for Group M(I) significant for either the first or second half, $T(30) = 224, z = .18, T(30) = 300, z = 1.30$, respectively.

TABLE II

MEDIAN THIRD-INTERVAL DIFFERENCE SCORES FOR TEST TRIALS DURING ADAPTATION AND ACQUISITION

Group	Accuracy of verbalization	N	Trial blocks: Adaptation 1–2	Acquisition 1–2	Acquisition 3–4
Masked (M)	Accurate M(A)	13	.02	−.11	−.05
	Inaccurate M(I)	30	−.04	.00	.03
	M(A) and M(I) combined	43	−.01	−.03	.00
Comparison (C)	Accurate C(A)	15	−.03	.24*	.24**
	Inaccurate C(I)	9	.03	.01	.04
	C(A) and C(I) combined	24	−.01	.18*	.10**

* $p < .05$.
** $p < .01$.

The median third-interval difference score shown in Table II for Group C(A) during the first half of acquisition was statistically significant, $T(15) = 24$, $p < .05$, while the score for Group C(I) was not, $T(9) = 12$. However, direct comparison of the differentiation for the C(A) and C(I) groups revealed no significant difference, $U(15, 9) = 51$. There was evidence that the two groups combined (Group C) differentiated reliably, $T(24) = 80$, $p < .05$. In the second half of acquisition, Group C(A) differentiated significantly, $T(15) = 9$, $p < .01$, while Group C(I) did not, $T(9) = 15$. It was also found that Group C(A) differentiated significantly more than Group C(I), $U(15, 9) = 20$, $p < .02$, and that the entire C group differentiated significantly, $T(24) = 56$, $p < .01$. The third-interval differentiation of the M and C groups was also compared. During the first half of acquisition, Group C differentiated significantly more than Group M, $U(24, 43) = 356$, $z = 2.09$, $p < .05$, though the two groups did not differ significantly in the second half, $U(24, 43) = 422$, $z = 1.23$.

DISCUSSION

Results of the postconditioning interview indicated that the masking procedure was effective in reducing the frequency with which *Ss* accurately verbalized the stimulus relations. While 62% of the *Ss* in Group C were classified as accurate verbalizers, only 30% of the *Ss* in Group M were so classified. The substantial proportion of accurate verbalizers in Group M emphasizes the importance of using a postconditioning interview or some other means to assess the cognitive effects of masking, rather than assuming its success. The effectiveness of masking may be expected to vary in different studies

due to various characteristics of the masking procedure, conditioning paradigm, and *S*s.

The expectation that there would be less differential conditioning in Group M than in Group C was strongly confirmed. Group M showed significantly less conditioning of first-interval GSRs during the combined acquisition trials as well as less conditioning of second-interval GSRs in both halves of the acquisition series. The inferiority of Group M in differentiation of third-interval GSRs was similarly apparent during the first half of acquisition. The reduced conditioning occurring in Group M is consistent with results obtained by Ross, Wilcox, and Mayer,[11] but the present results are not in accord with Razran's contention that simple salivary conditioning is facilitated by masking.[12] It should be noted, however, that Razran's groups contained only three *S*s each.

In Group C there was consistent evidence that differential GSR conditioning was limited to Group C(A), which contained *S*s who accurately verbalized the stimulus contingencies in the postconditioning interview. The GSR magnitudes of Group C(A) were significantly differentiated in the first interval during the combined acquisition trials, in the second interval during the last half of acquisition trials, and in the third interval during both halves of acquisition. The inaccurately verbalizing *S*s comprising Group C(I) did not differentially condition in any GSR interval. The concordance obtained between knowledge of stimulus relations and GSR differentiation is consistent with results of our previous studies using both intertrial reports and postconditioning interviews.[13] In addition, the present results for both the C(A) and C(I) groups replicate the outcome of a previous study of differential GSR conditioning using the same CS-UCS interval and only a postconditioning interview.[14]

In Group M, the relationship between conditional GSR differentiation and accuracy of verbalization was different for first- and second-interval responses. During the late acquisition trials, significant first-interval differentiation was observed for both the accurately verbalizing *S*s of Group M(A) and the inaccurately verbalizing *S*s of Group M(I). On the other hand, conditional differen-

[11] Ross, Wilcox, and Mayer, *loc. cit.*
[12] G. Razran, Attitudinal determinants of conditioning and of generalization of conditioning, *J. exp. Psychol.*, 30, 1949, 820–829.
[13] Fuhrer and Baer, *op. cit.*, *Science*, 150, 1965, 1479–1481; Baer and Fuhrer, *op. cit.*, *J. exp. Psychol.*, 78, 1968, 81–88.
[14] Fuhrer and Baer, *op. cit.*, *Proc. 74th Annu. Conv. Amer. Psychol. Ass.*, 1966, 53–54.

tiation of second-interval GSRs was consistent with the accuracy of postconditioning verbalization since Group M(A) differentiated reliably during the second half of acquisition and Group M(I) did not. In addition, the differentiation of Group M(A) was significantly greater than that of Group M(I). Similar comparisons for third-interval GSRs were not possible since the two groups did not show significant differentiation when combined or considered separately.

Razran has made a frequently noted distinction between relational learning and conditioning: relational learning refers to physiological evidence for conditioning in conjunction with reported perception of CS-UCS or CS-CR relations, and conditioning per se is defined as occurring in the absence of such perception.[15] Interpreting the present data in these terms, it appears that differential GSR conditioning obtained with the conventional procedure is almost exclusively a reflection of relational learning. On the other hand, the masking procedure appears to yield relational learning in some *S*s and conditioning in others. The latter findings contrast with the report by Dawson and Grings that conditioning does not occur when a masking task is used, *i.e.* that there is no evidence of GSR differentiation for *S*s unable to recognize the stimulus relations.[16] These investigators acknowledge, however, that their masking task may have had the effect of interfering with perception of the CSs, a difficulty which the present study avoided by prominently involving the CSs in the masking task itself. Razran's terminology also serves to highlight a potentially important distinction between first- and second-interval GSR differentiation. While first-interval differentiation may be observed in conjunction with both conditioning and relational learning, second-interval differentiation appears to be exclusively a correlate of relational learning. This formulation is consistent with the view that second-interval GSRs are emitted as concomitants of a preparatory set in anticipation of the UCS. It is probable that this set is under covert verbal control, particularly with respect to time estimation of the CS-UCS interval. On the other hand, first-interval GSRs may be viewed as conditional orienting responses which are much more under stimulus control and potentially more independent of verbal processes.

The differential GSR conditioning which occurred for Group M(I) may, alternatively, indicate that masking selectively reduced

[15] G. Razran, Conditioning and perception, *Psychol. Rev.*, 62, 1955, 83–95.
[16] Dawson and Grings, *loc. cit.*

the effectiveness of the interview to reflect *Ss'* covert verbal processes during conditioning. Some of the *Ss* who showed first-interval differentiation may have achieved at least a tentative formulation of the actual stimulus relations. However, attention to the masking task may have competed with the opportunity for these *Ss* to confirm and rehearse their formulation, and these *Ss* may have consequently lacked the subjective confidence necessary for verbalization during the interview. Supporting these conjectures is the fact that the less specific interview questions required recall rather than mere recognition of the correct stimulus relations. These considerations suggest that intertrial verbal reports may more sensitively reflect covert verbal processes occurring during a masked conditioning procedure. Intertrial reports place less burden on the *S's* memory, and they require a shorter and presumably less complex span of experience to be communicated.

Finally, it may be noted that certain aspects of the present results support either of two possible interpretations of the role of cognitive processes in the differential conditioning of intellectually capable human *Ss*. A view that conditioning is a consequence of cognitive processing of stimulus relations is supported by the severely debilitating effects of masking on conditioning. However, the alternative position, which precludes the necessity of cognition for conditioning, is also demonstrated, for under the unique circumstances of masking, *Ss* unable to verbalize the stimulus contingencies nevertheless showed some conditioning. A question for future research is whether experiments can be devised to adequately test these alternative viewpoints. It may be added that the relationship between conditioning and cognition observed in the comparison group was essentially correlative in nature and thus did not constitute a test of either viewpoint.

Summary

Differential GSR conditioning was assessed in a group ($N = 43$) which had the conditioning procedure embedded in a probability-learning task (Group M) and in a comparison group ($N = 24$) which underwent a conventional differential-conditioning procedure (Group C). The CSs were 8-sec. tones of differing frequencies, the UCS was a shock, and the CS-UCS interval was 8 sec. Division of the CS-UCS and the post-UCS period into three 4-sec. intervals permitted the analysis of multiple GSRs. Content analyses of post-conditioning interviews indicated that 62% of the *Ss* in Group C

verbalized the stimulus relations accurately, while only 30% of the *Ss* in Group M did so. As expected from a cognitive viewpoint, Group M showed significantly less conditional differentiation of GSRs in all three intervals. Significant GSR conditioning was observed in the three intervals for the 15 accurate verbalizers in Group C, but there was no evidence of conditioning in any interval for the 9 inaccurate verbalizers. The 13 accurate verbalizers in Group M showed significant second-interval differentiation while the 30 inaccurate verbalizers did not. However, both accurate and inaccurate verbalizers in Group M showed significant first-interval conditioning during the late acquisition trials.

NONVERBAL LEARNING

By B. G. Rosenberg and Jonas Langer,
University of California, Berkeley,
and John Stewart, Bowling Green State University

Students of the origins of language (*e.g.* Cassirer, Piaget, Werner and Kaplan)[1] have long recognized that nonverbal, particularly gestural and vocal, symbolization may play (1) an important developmental role in mediating the ontogenesis of conventional verbal symbolization and (2) an important communication role in supporting or contradicting the message in verbal statements with which it may be associated. The fundamental causal question about such nonverbal symbolization is whether it is a culture-bound (learned) or universal (natural) phenomenon.[2] This question of causation has been the theoretical impetus for numerous empirical studies, most of them matching experiments in which samples of *S*s were asked to match symbolic patterns in one medium with referents of another medium. The purpose has been to determine whether beyond-chance agreement obtains between the matchings of the *S*s in the sample. Two symbolic media, the sound and the graphic, have been studied fairly intensively in such matching experiments. In the medium of sound, *S*s have matched nonverbal but vocal sounds such as TAKETE and ULOOMU with visual patterns,[3] nonverbal sounds with dimensions of verbal meaning,[4] and antonymic pairs of words from historically unrelated languages.[5] In the graphic medium, *S*s have

* Received for publication October 4, 1968. The study was aided in part by a grant to the first author from the National Institute of Mental Health, MH-07994-05.

[1] E. Cassirer, *The Philosophy of Symbolic Forms*, Vol. I, *Language*, 1953; J. Piaget, *Play, Dreams, and Imitation*, 1951; H. Werner and B. Kaplan, *Symbol Formation*, 1963.

[2] See J. Langer and B. G. Rosenberg, Symbolic meaning and color naming, *J. pers. soc. Psychol.*, 4, 1966, 364–373.

[3] W. Kohler, Psychological remarks on some questions of anthropology, this Journal, 50, 1937, 271–288; R. Davis, The fitness of names to drawings: A cross-cultural study in Tanganyika, *Brit. J. Psychol.*, 52, 1961, 259–268.

[4] J. Langer and B. G. Rosenberg, The representation of verbal referents in sonic vehicles, *Percept. mot. Skills*, 19, 1964, 663–670; B. G. Rosenberg, P. Badia, and J. Langer, A representational assessment of meaningfulness, *Psychon. Sci.*, 1, 1964, 263–264.

[5] J. Atzet and H. B. Gerard, A study of phonetic symbolism among native

matched abstract linear patterns with dimensions of verbal meaning[6] and postural-gestural linear patterns with dimensions of verbal meaning.[7]

Nevertheless, very little work has been undertaken to demonstrate the compelling effect of nonverbal symbols when the idea of forming representational matchings is not directly suggested to the *S*s, as it has been in the matching studies referred to above. Since this suggestion is inherent in the matching technique itself, experimental demonstrations that do not use the matching technique are a theoretical necessity if the hypothesis of natural symbolization is to be sustained. Two studies have begun to demonstrate the coercive influence of nonverbal symbols in nonmatching situations. In the first study, it was found that the meaning intrinsic to nonverbal sound symbols may affect the maintenance of directed cognitive activity in much the same fashion as does the meaning of verbal symbols.[8] In the second, it was found that congruent pairs of sound symbols and verbal referents were learned more easily than the same sound symbols paired with noncongruent or unrelated referents.[9]

The present investigation was designed to be a further empirical step in the authors' systematic attempt to explore the natural coercive influences of the meaning intrinsic to a variety of nonconventional and nonverbal symbols when a matching technique is not employed. To this end, two experiments were conducted; both employed a learning paradigm. Experiment I dealt with graphic symbols and verbal referents. In a previous matching study, it was found that *S*s agreed in their matching of postural-gestural symbols with particular verbal referents (*e.g.* red, green, yellow, blue) beyond chance expectation.[10] These positive findings were interpreted as evidence supporting the contention that there is an intrinsic semantic relationship between certain postural-gestural symbols and verbal referents, *i.e.* that they are not extrinsically related, that they are not overtly or structurally similar, and

---

Navajo speakers, unpublished MS, 1965; R. Brown and R. Nuttal, Methods in phonetic symbolism experiments, *J. abnorm. soc. Psychol.*, 59, 1959, 441–445.

[6] M. Scheerer and J. Lyons, Line drawings and matching responses to words, *J. Pers.*, 25, 1957. 251–273; Werner and Kaplan, *loc. cit.*

[7] B. G. Rosenberg and J. Langer, A study of postural-gestural communication, *J. pers. soc. Psychol.*, 2, 1965. 593–597.

[8] Langer and Rosenberg, *op. cit.*, *J. pers. soc. Psychol.*, 4, 1966, 364–373.

[9] J. Langer, E. E. Sampson, and B. G. Rosenberg, Learning verbal referents of phonetic symbols, *J. pers. soc. Psychol.*, 3, 1966, 427–434.

[10] Rosenberg and Langer, *loc. cit.*

that they are not linked by arbitrary convention. Yet, *S*s must 'perceive' some inner (natural physiognomic) relationship for the dematching agreement to obtain. Experiment I was therefore designed to determine whether the intrinsic semantic relationship between gestural symbols and their referents directly influences paired-associate learning (as was the case with the nonverbal sound symbols mentioned above).

Experiment II was a first attempt to study, in a nonmatching experiment, the effect of nonverbal symbolism upon an aspect of cognition when not directly mediated by verbal referents. To date, as mentioned above, it has been possible to demonstrate the effect of the semantic relationship between nonconventional symbols and verbal referents on aspects of cognitive behavior. In Experiment II, the attempt was to demonstrate the effect of the intrinsic semantic relationship between symbols in two different nonverbal media upon paired-associate learning. To this end, *S*s were presented with semantically congruent and semantically noncongruent lists of gestural symbols paired with sound symbols for learning, the gestural[11] and sound[12] symbols having been previously found to be consensually matched with the same or different color referents by independent samples of *S*s.

Experiment I

Graphic Symbols and Verbal Referents

Lists. Two unmixed lists of 10 paired associates were constructed. One list contained congruent and the other noncongruent pairs. The stimuli in the pairs were stick figures,[13] while the responses were chromatic or achromatic color names. The stimulus and response items were selected on the basis of previous findings on consistency in matching stick figures with colors.[14] *Congruent* pairs were defined as those for which high consistency in matching stick figures to colors had been obtained. The two highest for red, yellow, black and white, and the one highest for blue and grey, were utilized. *Noncongruent* pairs were formed by randomly assigning each figure to one of the other two chromatic or achromatic colors, but with the restriction that there be two red, two yellow, two black, two white, one blue, and one grey. Thus, each list consisted of 10 pairs. The set of stick-figure stimuli was the same for both lists. Each color-name response also appeared with equal frequency on both lists. On the congruent list each stick figure was paired with the color most consistently matched with it, while on the noncongruent list the figure was paired with the

[11] Rosenberg and Langer, *loc. cit.*
[12] Langer and Rosenberg, *op. cit.*, *Percept. mot. Skills*, 19, 1964, 663–670.
[13] T. R. Sarbin and C. Hardyck, Conformance in role perception as a personality variable, *J. consult. Psychol.*, 19, 1955, 109–111.
[14] Rosenberg and Langer, *loc. cit.*

TABLE I

STIMULUS AND RESPONSE MATERIALS: EXPERIMENT I

Number	Stick figure (stimulus)	Percentage of agreement	Congruent color (response)	Noncongruent color (response)
34		75	Red	Blue
35		79	Red	Yellow
18		51	Yellow	Red
50		47	Blue	Yellow
44		46	Yellow	Red
23		97	White	Black
19		83	Black	White
38		80	Black	Grey
25		68	White	Black
22		57	Grey	White

noncongruent color assigned to it. The stimulus figures, congruent and noncongruent responses, and percentage matching agreement[15] are presented in Table I. To minimize the opportunity for serial learning, the lists were prepared in four random orders. Randomization was restricted so that no pair was permitted to follow itself for at least five pairs.

Procedure. The lists were presented on a Stowe memory drum set to operate at a 3:3-sec. rate. The *Ss* were told that they would be shown some stick figures of people in various postures and that each figure would be paired

[15] Rosenberg and Langer, *loc. cit.*

with the name of a color. On the first presentation, *S* was instructed to read each color name aloud. *Ss* were then given the standard paired-associate instructions for the anticipation method with guessing encouraged. All *Ss* were required to learn the list to a criterion of one correct recitation.

Subjects. Thirty undergraduate students at Bowling Green State University served as *Ss*. Fifteen *Ss* (6 men and 9 women) were tested on the congruent list; another 15 (6 men and 9 women) were tested on the noncongruent list All *Ss* were volunteers.

Results. Fig. 1 presents the mean number of correct responses per trial for the two conditions. All *Ss* are represented at each point, with a perfect score assigned to *S* on trials after he reached criterion. The curves are terminated at the point at which all *Ss* reached criterion. Clearly, the congruent gesture-color pairs were learned more readily than the noncongruent pairs.

Table II presents the main results. Since Jenkins has indicated that reliable differences observed in verbal-learning experiments are largely found in the earlier phases of list learning,[16] and analysis of the number correct in the first two trials is also presented. Some *Ss* on the congruent list had reached criterion by the second trial; therefore differences in the number correct are presented for only the first two trials. As can be seen, whether scoring by number correct, overt errors per trial, or trials to criterion, congruent pairs were learned more rapidly than noncongruent pairs.

TABLE II

MEANS AND t VALUES FOR TRIALS TO CRITERION, OVERT ERRORS, AND NUMBER CORRECT IN FIRST TWO TRIALS: EXPERIMENT I

Measure	Congruent	Noncongruent	t value
Trials to criterion	7.00	13.33	3.370***
Overt errors per trial	2.35	3.17	2.216**
Number correct in first two trials	10.60	7.27	2.408**

** $p < .02$, one-tailed.
*** $p < .005$, one-tailed.

EXPERIMENT II

GESTURAL SYMBOLS AND SOUND SYMBOLS

Lists. Two unmixed lists of eight paired associates were constructed. One list contained congruent and the other noncongruent pairs. The stimuli in the pairs were stick figures, while the responses were sound patterns. The stimulus and the response items selected were those which had been consistently matched with the same colors by independent groups of *Ss*.[17] *Congruent* pairs were defined as pairs composed of gestural and sound symbols which share com-

[16] J. J. Jenkins, Comments on pseudomediation, *Psychon. Sci.*, 2, 1965, 97–98.
[17] Rosenberg and Langer, *loc. cit.*; Langer and Rosenberg, *op. cit.*, *Percept. mot. Skills*, 19, 1964, 663–670.

mon color referents. That is, a given pair was made up of (1) a stick figure which had been matched with high consistency to a particular color and (2) a sound pattern that had also been consistently matched with the same color. For each of the colors (red, yellow, green, and blue), the two stick figures and the two sound patterns of highest consistency were selected and randomly picked. *Noncongruent* pairs were formed by randomly assigning each stick figure to a sound pattern that did not share the same color referent. However, the same sets of figures and sounds were used. Thus, each list consisted

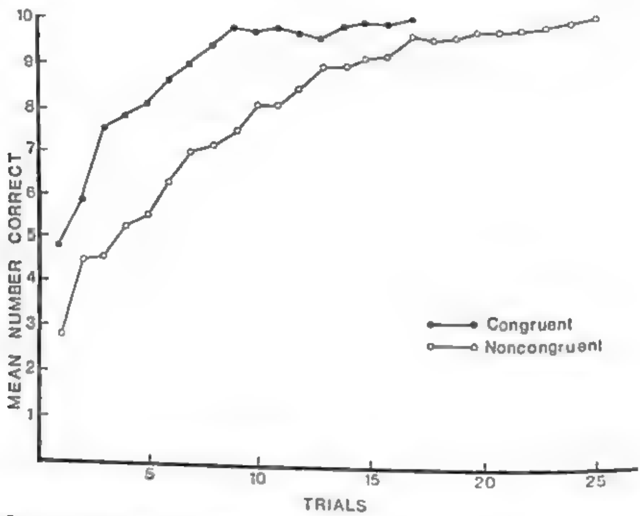


FIG. 1. MEAN NUMBER CORRECT RESPONSES ACROSS TRIALS: EXPERIMENT I

of eight pairs. The stick-figure stimuli were the same in both lists. In addition, the same set of sound responses was used. Consequently, the sound responses also appeared with equal frequency on both lists. The two lists are presented in Table III. In order to minimize the opportunity for serial learning, the lists were prepared in five random orders. Randomization was restricted so that no pair was permitted to follow itself for at least four pairs.

Procedure. The lists were presented by the anticipation method on a Stowe memory drum set to operate at 3:3-sec. rate. *Ss* were told that they would be shown some stick figures of people in various postures and that each figure would be paired with a sound pattern. On the first presentation, *Ss* read each sound aloud. They were then given the standard paired-associate instructions for the anticipation method with guessing encouraged. All *Ss* were required to learn the list to a criterion of one correct recitation.

Subjects. Thirty undergraduate students at Bowling Green State University served as *Ss*. Fifteen *Ss* (5 men, 10 women) were tested on the congruent list, and another 15 (6 men, 9 women) were tested on the noncongruent list. All *Ss* were volunteers.

Results. Fig. 2 presents the mean number of correct responses by trial. Values for plotting were determined in the same manner

TABLE III

Stimulus and Response Materials: Experiment II

Number	Stick figure (stimulus)	Color referent	Congruent sound (response)	Noncongruent sound (response)
16		Green	Nerd	Klak (red)
18		Yellow	Elbe	Com (blue)
31		Blue	Mumle	Zah (red)
34		Red	Klak	Nerd (green)
35		Red	Zah	Elbe (yellow)
39		Blue	Oom	Zing (yellow)
44		Yellow	Zing	Tur (green)
46		Green	Tur	Mumle (blue)

Note: Colors in parentheses are the referents of the noncongruent sounds.

as in Experiment I. As in Experiment I, such plotting provided nonoverlapping curves. Clearly, congruent symbol pairs were more readily learned than noncongruent symbol pairs.

Table IV presents the mean trials to criterion, overt errors per trial, and the number correct in the first five trials. As is evident from the means, the results were as predicted, though the level of acceptability was not as significant as in Experiment I. Again, whether scoring by trials, errors, or number correct, stick figure–sound pattern congruencies were more readily learned than stick figure–sound pattern noncongruencies.

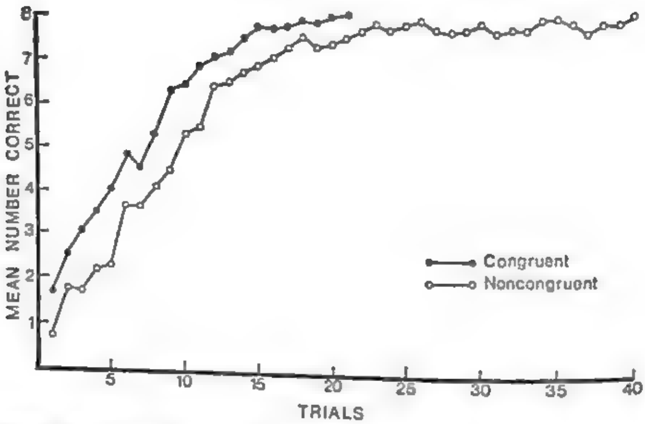


Fig. 2. Mean Number Correct Responses across Trials: Experiment II

TABLE IV

Means and t Values for Trials to Criterion, Overt Errors, and Number Correct in First Five Trials: Experiment II

Measure	Congruent	Noncongruent	t value
Trials to criterion	11.87	15.73	1.650*
Overt errors per trial	2.42	2.99	1.781**
Number correct in first five trials	14.60	8.60	3.371***

* $p < .10$, one-tailed.
** $p < .05$, one-tailed.
*** $p < .01$, one-tailed.

Discussion

At best, matching experiments permit demonstration that there is an intrinsic semantic relationship between nonverbal symbols and referents when the representational relationship is directly suggested to *S* by the experimental method itself.[18] In two previous studies, the present authors have found that the hypothetical, intrinsic semantic relationship between nonverbal symbols and referents has a demonstrable effect even when the representational relationships are not suggested to *S*. Methodologically, this has meant demonstrating the effect of the hypothetical relationships on some aspects of cognition in situations which do not involve matching. The technique, then, does not capitalize upon the suggestive qualities inherent in matching experiments. Rather, it is one step removed from the matching situation so that it does not itself suggest the hypothetical relationships that it is presumed to account for. Using this technique, the present study further

[18] For a fuller discussion of this methodological critique, see Brown, *Words and Things*, 1958, and Langer and Rosenberg, *op. cit.*, *J. pers. soc. Psychol.*, 4, 1966, 364–373.

bolsters the empirical basis for maintaining that there are indeed certain intrinsic semantic relationships between nonverbal symbols (here, postural-gestural symbols) and verbal referents (here, colors). The evidence comes from Experiment I, where it was found that congruent pairs of postural-gestural symbols and verbal referents were learned more easily than noncongruent pairs.

From the methodological-empirical point of view, however, the results of Experiment II are of even greater import. In that experiment it was possible to demonstrate, for the first time, the coercive effect of the hypothetical, intrinsic semantic relationships when the situation is even more removed from the suggestive character of matching experiments. To begin with, *S*s did not directly match nonverbal symbols with referents. More important, they did not even learn pairs of nonverbal symbols and referents. Rather, they learned pairs of nonverbal symbols: postural-gestural with sound symbols. Experiment II is therefore particularly relevant to the theoretical question of the cause of nonverbal symbolization. *S*s can hardly be presumed to have experienced these particular pairs of postural-gestural and sound symbols together before this experiment. Yet, they learned congruent pairs much more readily than noncongruent pairs. To begin to explain these results, it seems necessary to posit that symbols in different media, such as the postural-gestural graphic and the phonetic medium, may share common spheres of meaning.[19] In the present situation, this would imply that when a given color is the referent of both a particular postural-gestural symbol and sound symbol, it may serve to mediate both.[20] The possible mediational value of the color is thus not restricted to when it is overtly present in the situation where the symbols are being considered. Rather, it may have a *covert* presence. It may be an implicit, but at least a partial, defining characteristic of the sphere of meaning shared by the two different kinds of symbols. It may be why congruent symbol-symbol pairs (in addition to congruent symbol-referent pairs) are learned more easily than noncongruent pairs. In sum, the present study reinforces the hypothesis that there are intrinsic semantic relationships shared by nonverbal symbols to account for successful matchings of symbols that have little in common extrinsically.

[19] See M. Scheerer, Spheres of meaning, *J. indiv. Psychol.*, 15, 1959, 50–61.
[20] The reason for qualifying this explanation is that we may be incorrect in identifying our interpretation with the experimental operations we employed.

Summary

Two studies were performed. In the first, it was found that congruent symbol-referent pairs of postural-gestural symbols and color referents were learned more readily than noncongruent pairs. In the second, it was found that congruent symbol-symbol pairs of postural-gestural symbols and sound symbols were learned more readily than noncongruent pairs. Taken together, these findings from nonmatching experiments support the notion of intrinsic semantic relationships that are not the result of learning arbitrary, conventional relationships.

THE REFERENCE FOR VISUAL NORMALIZATION

By R. H. Day and N. J. Wade,
Monash University, Australia

Prolonged inspection of a tilted line results in an apparent decrease in the tilt of that line, an effect referred to as normalization.[1] Normalization has been measured indirectly from the difference between judgments of the visual vertical before and after a period of inspection, an effect referred to as the visual tilt aftereffect. Following inspection of a tilted line, the visual vertical is shifted in the direction of prior tilt or, alternatively, a gravitationally vertical line is perceived as tilted in the direction opposite to that of the inspection line. Gibson considered that this tilt adaptation tended in the direction of some perceptual norm, although he did not specify its nature.[2] Experiments concerned with such visual tilt aftereffects have generally involved *O*s in the normal upright posture, *i.e.* with the head (and, therefore, the retinal meridian) in the gravitational vertical. It may be argued that under such conditions normalization occurs relative to the retinal meridian or to gravity. The purpose of the experiments reported here was to establish the reference relative to which normalization occurs.

The retinal and gravitational reference axes may be separated by laterally tilting the head or whole body, a method used previously in establishing the reference axis for shape recognition and the basis of deterioration in visual acuity with object tilt.[3] From the viewpoint of the present problem, a gravitationally vertical line with body tilted would be tilted relative to the retinal meridian and, contrariwise, a line in the retinal meridian would be tilted relative to gravity. If normalization occurs relative to the retinal merid-

* Received for publication October 21, 1968.

[1] J. J. Gibson, Adaptation, aftereffect and contrast in the perception of curved lines, *J. exp. Psychol.*, 16, 133, 1–31; J. J. Gibson and M. Radner, Adaptation, aftereffect and contrast in the perception of titled lines: I, Quantitative studies, *J. exp. Psychol.*, 20, 1937, 453–467.

[2] Gibson, *loc. cit.*

[3] I. Rock, The orientation of forms on the retina and in the environment, this Journal, 69, 1956, 513–528; I. Rock and W. Heimer, The effect of retinal and phenomenal orientation on the perception of form, this Journal, 70, 1957, 493–511; S. M. Luria, The effect of body-position on meridional variation in scotopic acuity, this Journal, 76, 1963, 598–606.

ian, an aftereffect would be expected from the former case but not from the latter, and the reverse would apply for normalization to the gravitational vertical. Data suggesting normalization to the gravitational vertical for a laterally tilted *O* were reported by Prentice and Beardslee, but the magnitude of the effect was not measured.[4] It is also worth noting that discriminative reaction times to vertical and horizontal lines are faster than those to oblique lines, and that this difference still obtains when the head is tilted 45°.[5]

Method

Conditions. In order to determine the reference axis for visual normalization the visual tilt aftereffect was examined under the four conditions given by two body postures and two line orientations in a preliminary experiment, and under the two critical conditions of these four in the main experiment. Two hypotheses, retinal and gravitational, were postulated. According to the retinal hypothesis, the inspection of a line inclined to the retinal meridian generates the visual tilt aftereffect. The gravitational hypothesis predicts that inclination of the inspection line to gravity produces the aftereffect. The four combinations of body posture and inspection-line orientation are shown diagrammatically in Fig. 1, together with the predictions from the retinal and gravitational hypotheses.

The task in all conditions involved adjustment of a line of light to the visual vertical before (pretest) and after (posttest) a 2-min. inspection period. In Condition I, with the body upright, the inspection figure was in both the gravitational vertical and retinal meridian, and therefore no visual tilt aftereffect would be expected. Under Condition II the inspection figure was tilted 15° clockwise of both the retinal and gravitational axes. This is the traditional tilt aftereffect condition, and both hypotheses predicted that the posttest visual vertical would be clockwise of that for the pretest. Inclination of the body 15° to the right in Condition III with the line in the gravitational vertical led to different predictions from the two hypotheses. The inspection figure was tilted 15° counterclockwise relative to the retinal meridian and, therefore, the retinal hypothesis predicted an aftereffect in the counterclockwise direction. No aftereffect was predicted from the gravitational hypothesis, since the inspection figure was in the gravitational vertical. Conversely, in Condition IV both the body and the line were tilted 15° clockwise and the line remained in the retinal meridian, ignoring ocular countertorsion.[6] No aftereffect was predicted from the retinal hypothesis, but a clockwise aftereffect was predicted from the gravitational one.

[4] W. C. H. Prentice and D. C. Beardslee, Visual "normalization" near the vertical and horizontal, *J. exp. Psychol.*, 40, 1950, 355–364.

[5] F. Attneave and R. K. Olson, Discriminability of stimuli varying in physical and retinal orientation. *J. exp. Psychol.*, 74, 1967, 149–157.

[6] During body tilt the eyes undergo ocular countertorsion, so that the retinal meridian would no longer lie in the head axis. For 15° body tilt, the countertorsion is approximately 2° (E. F. Miller, Counterrolling of the human eyes produced by head tilt with respect to gravity, *Acta Otolaryng.*, 54, 1961, 479–501; H. Schöne, Ueber den Einfluss der Schwerkraft auf die Augenrollung und auf die Wahrnehmung der Lage im Raum, *Z. vergl. Physiol.*, 46, 1962, 57–87.

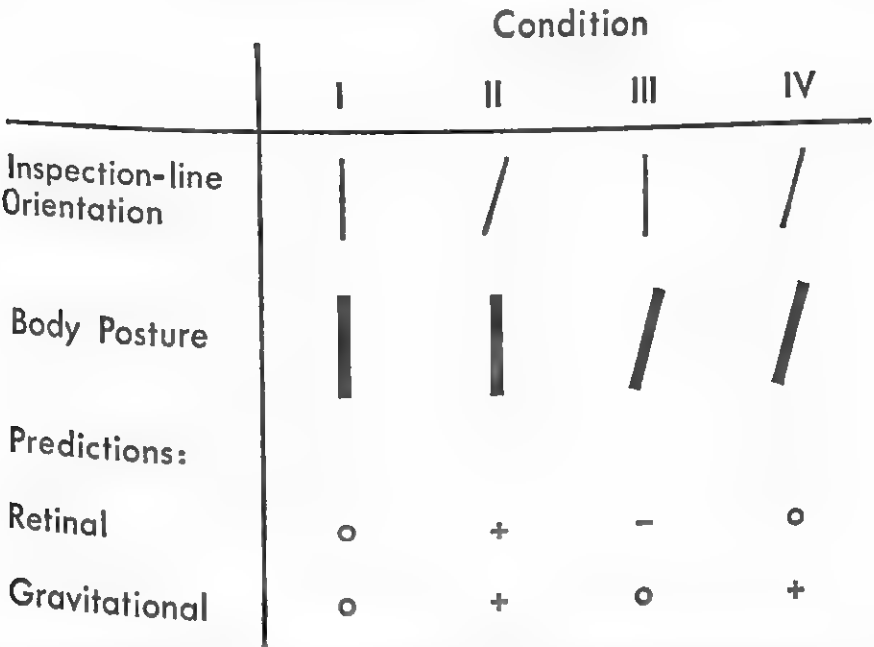


FIG. 1. INSPECTION-LINE ORIENTATIONS AND BODY POSTURES, WITH PREDICTED AFTEREFFECT DISPLACEMENTS FROM RETINAL AND GRAVITATIONAL HYPOTHESES (+ indicates clockwise; — indicates counterclockwise, 0 indicates no prediction.)

Conditions III and IV were, of course, the only ones which permitted direct comparison of the two hypotheses. However, Conditions I and II were included in the pilot experiment (1) so that some indication of the magnitude of the aftereffect under the particular conditions could be assessed and (2) so that the aftereffect under Condition III could be compared with that of Condition II. Both II and III involved line tilt relative to the retinal meridian. Only Conditions III and IV were included in the main experiment.

Procedure. The apparatus for controlling body posture has been described elsewhere.[7] Adjustable leg, hip, and shoulder supports minimized any trunk movement, and head position was controlled by side supports and a dental bite board. The visual stimulus object was a line 78 mm. long and 0.3 mm. wide, corresponding to an aperture in an otherwise lighttight lamp housing, 120 cm. from *O*'s eyes and with a luminance of 0.08 ftL. The line could rotate frontoparallel to *O* with its center in the median plane and adjustable to eye level. The orientation of the line, which could be measured to the nearest 0.1°, was controlled either by *O*, using a switch with his right hand, or manually by *E*. The switch activated a synchronous 1-rpm motor. The lamp housing was fixed to the box, and the whole arrangement could be laterally tilted while maintaining the same relation between the stimulus object and *O*. The apparatus was housed in a dark room.

[7] N. J. Wade, Visual orientation during and after lateral head, body, and trunk tilt, *Percept. & Psychophys.*, 3, 1968, 215–219.

Three *O*s, well-practiced in making verticality judgments under the conditions used, participated in the pilot study. Twenty unpracticed *O*s took part in the main experiment. Each *O* underwent four aftereffect trials in a session. A trial consisted of four pretest adjustments to the visual vertical, one each from starting positions of 5° and 10° clockwise (+) and counterclockwise (−) of the gravitational vertical, presented in random order. The *O*'s eyes were closed at all times other than when making an adjustment. There followed a 2-min. inspection period with the line and body in the appropriate orientations, after which a single posttest adjustment was made from the first of the pretest starting positions. The interval between cessation of inspection and the initiation of the posttest was several seconds, during which time *O*'s eyes were closed. The difference between the posttest and the mean of the pretest judgments was the measure of the visual aftereffect. A 3-min. intertrial interval was allowed, during which *O* remained upright with eyes closed.

The three *O*s in the preliminary study took part in four sessions, each of which consisted of four trials, one under each of the four conditions which were presented in random order. The 20 *O*s in the main experiment underwent two trials, one for each condition, during a single session. The order of these two conditions was randomized.

Results

In the pilot study the mean aftereffects for Conditions I, II, III, and IV were −0.30, +1.44, +0.73, and +2.50 respectively. This pattern of aftereffect magnitudes, which was general for the three *O*s, favors the gravitational hypothesis. However, the small number of *O*s and relatively large variances may be criticized. Furthermore, all *O*s were trained and more or less sophisticated in regard to the problem. For these reasons, it is desirable to consider in detail only the data from the main experiment.

For the main experiment the mean aftereffect with line vertical and body tilted (Condition III) was +0.64; with line and body tilted (Condition IV), it was +2.27. The data were analyzed by means of t tests on the pre- and posttest differences for each condition separately. The null hypothesis was rejected for $t > 2.093$ ($df = 19$, $\alpha = 0.05$). The value of t for Condition III was 1.28 and for Condition IV 4.61. Thus the mean aftereffect under Condition III was slightly clockwise but nonsignificant, and that under Condition IV clockwise and significant.

Discussion

The data from the pilot experiment suggested that visual normalization occurs relative to the gravitational reference rather than to the retinal. While the aftereffects were large and in the expected direction for Conditions II and IV, in which the line was tilted

relative to the gravitational axis, the aftereffect for Condition III, in which the line was retinally tilted but gravitationally vertical, was markedly smaller. The results from the main experiment confirmed the suggestion that normalization is relative to the gravitational axis. Under Condition III no significant aftereffect occurred, but under Condition IV there was a larger and significant aftereffect.

Although these results strongly favor the gravitational hypothesis, several aspects of obtaining visual-orientation judgments during body tilt may be noted. (1) Ocular countertorsion can produce a slight difference between the retinal meridian and body tilt. However, countertorsion would have affected only the retinal hypothesis, which was rendered untenable by the experiments. Prentice and Beardslee did find a visual tilt aftereffect when the retinal meridian was aligned with the tilted inspection line.[8] Their experiment controlled for any countertorsion by generating an afterimage of vertical lines with the head upright and then tilting the head to the position where the afterimage was aligned with the inspection line. It would appear, therefore, that the countertorsion produced by tilt of the body is of minimal importance in the aftereffect judgments during tilt. (2) The visual vertical with 15° right tilt was slightly counterclockwise of the gravitational vertical referred to as the E effect.[9] Therefore, the gravitationally vertical inspection line would not have been visually vertical in Condition III. This factor is of importance only if the visual axis, rather than the gravitational axis, is the reference of normalization. This is discussed in more detail below.

(3) It is possible that proprioceptive adaptation, which would develop during prolonged tilt in Conditions III and IV, influenced visual orientation independently of the visual-inspection conditions. The visual vertical with body upright is influenced by a period of prolonged body tilt with eyes closed.[10] This visual aftereffect from prolonged postural stimulation has been attributed to adaptation of the proprioceptive receptors which signal the position

[8] Prentice and Beardslee, *loc. cit.*

[9] M. Bauermeister, Effect of body tilt on apparent verticality, apparent body position, and their relation, *J. exp. Psychol.*, 67, 1964, 142–147; E. F. Miller, A. R. Fregly, G. van den Brink, and A. Graybiel, Visual localization of the horizontal as a function of body tilt up to ±90° from gravitational vertical, NSAM-942, NASA Order No. R-47, Pensacola, Fla.: Naval School of Aviation Medicine, 1965; Wade, *loc. cit.*

[10] R. H. Day and N. J. Wade, Visual spatial aftereffect from prolonged head-tilt, *Science*, 154, 1966, 1201–1202.

of the body relative to gravity.[11] It may then be postulated that the proprioceptive adaptation which occurs during prolonged body tilt would similarly influence judgments during tilt; *i.e.* judgments made at the beginning of the tilt period would differ from those made some time later. This view is supported by the similarity of the difference between Conditions I and III (+1.03°) to that between Conditions II and IV (+1.06°) in the pilot study. The only factor that was the same for both body-tilt conditions was the period of tilt, and this would have influenced the judgments in the same (clockwise) direction independently of visual stimulation. Thus, it would appear that proprioceptive adaptation during prolonged tilt produced the greater visual aftereffect in Condition IV relative to Condition II.

The demonstration that visual normalization occurs relative to the gravitational reference rather than to the retinal reference casts doubt on theories of visual tilt aftereffects which deal with the visual system alone. Since normalization occurs relative to gravity, irrespective of body position, there must be postural information for the inclination of the body to gravity, and this must interact with that from the visual system. That such information for body posture can be utilized has been clearly shown from experiments concerned with visual orientation during tilt.[12] The interaction of the visual with the postural system is entirely consistent with the data on the visual aftereffect following prolonged head or body tilt.[13] Visual orientation with head upright is systematically influenced by prolonged tilt such that the visual vertical is shifted in the direction of previous tilt. Gibson's theory of adaptation to a perceptual norm may be able to handle these data in broad terms, although there was no attempt to specify this perceptual norm. The present experiments indicate that this norm may be gravity, although there exists the possibility that it may be the visual vertical. Since the visual vertical is in close correspondence to the gravitational vertical during tilt, the results are consistent with this interpretation.

There is some indirect evidence that normalization occurs to the visual rather than the gravitational vertical. The faster discriminative reaction times to vertical lines, as opposed to oblique lines, was

[11] R. H. Day and N. J. Wade, Involvement of neck proprioceptive system in visual after-effect from prolonged head tilt, *Quart. J. exp. Psychol.*, 20, 1968, 290–293; Wade, *loc. cit.*

[12] Bauermeister, *loc. cit.*; Miller, Fregly, van den Brink, and Graybiel, *loc. cit.*; Wade, *loc. cit.*

[13] Day and Wade, *op cit.*, *Science*, 154, 1966, 1201–1202.

not as pronounced with head tilted 45°.[14] Under these conditions an E effect would have been generated; *i.e.* the visual and gravitational verticals would have been slightly separated.[15] This difference in reaction times may have been due to the apparent tilt of the gravitationally vertical line.

Summary

Two experiments were conducted to investigate whether visual normalization occurs in the direction of the gravitational vertical or the retinal meridian. These latter were separated by measuring the visual tilt aftereffect during lateral body tilt. Conditions under which the inspection line was gravitationally vertical but tilted relative to the retinal meridian, and in the retinal meridian but tilted relative to gravity, generated opposite predictions from the gravitational and retinal hypotheses. The results from both experiments indicate that normalization occurs toward the gravitational reference. The possible effects of ocular countertorsion, the E effect, and postural adaptation were discussed.

[14] Attneave and Olson, *loc. cit.*
[15] Wade, *loc. cit.*

VERBAL-DISCRIMINATION LEARNING WITH VARYING NUMBERS OF RIGHT AND WRONG TERMS

By BENTON J. UNDERWOOD and JOEL S. FREUND,
Northwestern University

When words are presented *S* in a verbal-learning task, several characteristics or attributes of each word may become a part of the memory for the word. One of these attributes is the frequency of presentation in the experimental situation, for it is known that judgments of differential frequency of presentation are quite accurate even when *S* is not anticipating that such judgments will be called for.[1] This ability to make rather fine discriminations of frequency provides a plausible basis for the assumption that in verbal-discrimination learning the detection of frequency differences is a basic mechanism in the acquisition.[2] In the usual verbal-discrimination task, pairs of words are presented, one word in each pair being arbitrarily designated as the correct or right (R) word, the other as the wrong (W) word. When presented for learning by the anticipation method, the pair is shown briefly followed by the appearance of the R word only. As a working assumption, it is presumed that as a consequence of these procedures the R word gains two frequency units on each trial; the W word, one unit. If this difference in frequency is sufficient for a discrimination, *S* will be correct in choosing the higher-frequency word after a single trial.

A test of one of the more counter-intuitive predictions which stems from the theory was reported in an earlier article.[3] This prediction stated that if one R word was paired with two different W words, learning would be more rapid than if one W word was paired with two different R words. Assuming the accretion of frequency units as noted above, in the former paradigm the ratio

* Received for publication January 9, 1969. This work was supported by Contract Nonr-1228(15), Project NR 154-057, between Northwestern University and the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose by the United States Government.

[1] B. J. Underwood, Some correlates of item repetition in free-recall learning, *J. verb. Learn. verb. Behav.*, 1969, in press.

[2] B. R. Ekstrand, W. P. Wallace, and B. J. Underwood, A frequency theory of verbal-discrimination learning, *Psychol. Rev.*, 73, 1966, 566–578.

[3] Ekstrand, Wallace, and Underwood, *op. cit.*, 566–578.

should be 4:1 in favor of the R word after a single trial, while in the latter paradigm both the W word and the R words should gain two units. In fact, if these values are correct, learning based upon a frequency difference should be impossible in the second paradigm. However, the basic notion of the theory was assumed to be supported when it was shown that the 1R-2W paradigm was a much easier task than that produced by the 1W-2R paradigm.

As a followup, the present study was a systematic exploration of the influence of varying numbers of R and W words. The list length here was always 12 pairs. In one set of four conditions, there were either 2, 4, 6, or 12 different R words, with the number of different W words constant at 12. The frequency theory clearly predicts that learning rate should be inversely related to number of R words. When there were only two R words, each was paired with six different W words. After a single study trial the frequency differential would be 12:1 in favor of the R words; learning should occur after a single study trial. As the number of R words increases from 2 to 12, the frequency differential between R and W words would decrease and the learning rate should be retarded correspondingly.

In a comparable set of four conditions, the number of different R words was held constant at 12, with the number of different W words being 2, 4, 6, or 12. With only two W words, the frequency differential for a pair after a single trial would be 6:2 in favor of the W word. If this difference were sufficient for a frequency discrimination, and if *S* followed a rule of choosing the least frequent member of the pair, learning should occur on a single trial. However, it can be seen that the frequency difference (2:6—here reversed for R:W comparison) was appreciably less than in the parallel 2R list (12:1). With four different W words, the R:W ratio was 2:3, with six W words, 2:2. Therefore, the difficulty in learning should increase as the number of W words increases from two to six, difficulty being maximal at six. With 12 different W words, the task should become easier, for with 12 W words and 12 R words (the typical task), the ratio after a single trial should be 2:1 in favor of the R word.

Method

Lists. All lists contained 12 pairs of moderate-frequency, two-syllable nouns. In List 2R, two different R words were used as correct words, each being paired with six different W words. Exactly the same list was used as List 2W except that the two words which were correct words in List 2R were incor-

rect in List 2W; thus, each W word was paired with six different R words. In List 4R, four different R words were used (the two appearing in List 2R plus two more), each being paired with three different W words. The roles of the words (*i.e.* whether correct or incorrect) were reversed in List 4W. List 6R contained six different R words (those four used in 4R plus two more), each paired with two different W words; for List 6W, the roles were reversed. Lists 12R and 12W were exactly the same, but with 12 of the words correct in one version, and the other 12 correct in the other version.

Procedure and subjects. Learning was by the anticipation method. The pair occurred together for 1.5 sec., after which the shutter of the memory drum lifted to expose the correct word alone for 0.5 sec. The intertrial interval was 2 sec. On the first trial *S* did not guess. Following this study trial, four anticipation trials were given, *S* being instructed to respond to each pair by calling out the word he believed correct. During the 1.5-sec. exposure period one word occurred above the other, these positions being varied randomly across trials.

A total of 20 *Ss* was assigned to each of the eight lists. Lists were randomized within each of 20 blocks, and *S* was assigned to a given list in order of his appearance at the laboratory. All *Ss* were college students.

Results

Since all *Ss* responded to each pair on each anticipation trial, correct responses and errors were reciprocal. For presentation here, mean errors are used as the response measure. A plot of mean errors on (anticipation) Trial 1 and a plot of mean total errors across Trials 1–4 showed essentially the same relationships. Mean errors on Trial 1 are plotted in Fig. 1.

Looking first at the results for the R lists, it can be seen that the mean numbers of errors increased directly as number of different R words increased. Of the 20 *Ss* given List 2R, 16 showed perfect performance on Trial 1; the remaining four *Ss* produced a total of five errors. A total of two errors occurred on Trial 2, none on Trials 3 and 4. Thus, a single study trial produced nearly perfect learning. On List 4R, eight *Ss* made at least one error on Trial 1; 10 *Ss* did so on Trial 1 for List 6R; and all 20 *Ss* made at least one error on List 12R.

On List 2W, 15 *Ss* made at least one error on Trial 1; and on Lists 4W and 6W, all *Ss* made at least one error on this trial. Performance on Lists 4W and 6W was better than chance (six errors), but this was not statistically better than chance for List 4W ($t = 1.67$). The mean number of errors for List 4W on Trial 4 was 1.75, and for List 6W, 1.90. Clearly, these two lists provided a difficult learning task in comparison with the other lists.

An overall statistical analysis showed that R vs. W, number of

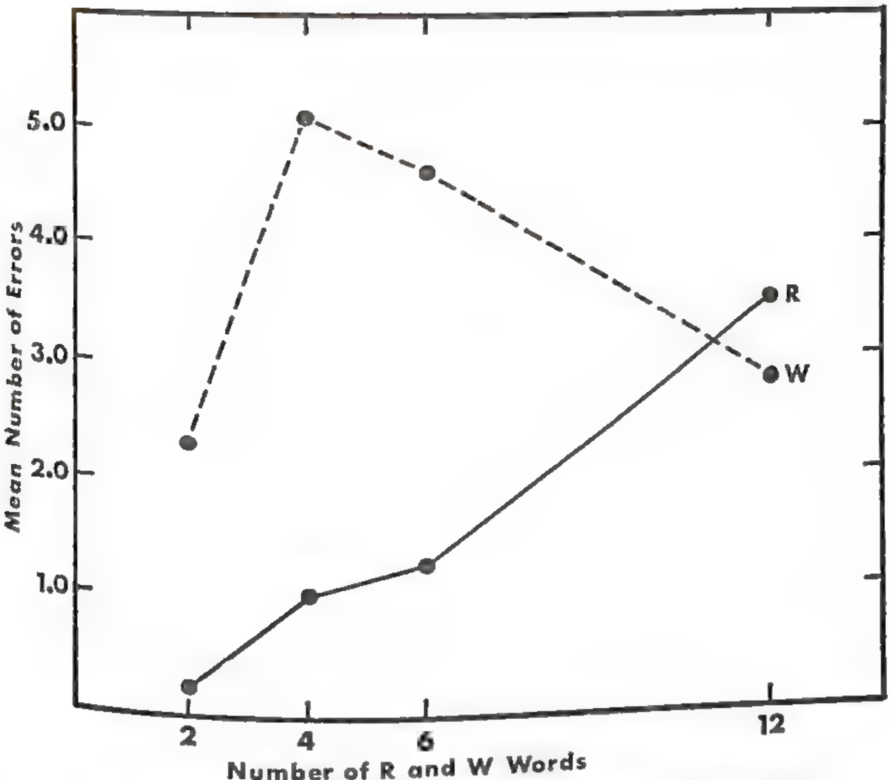


FIG. 1. MEAN ERRORS IN LEARNING AS A FUNCTION OF NUMBER OF R AND W WORDS

R and W words, and the interaction of these two variables were highly significant statistically. However, with so many zero scores for the R lists, the appropriateness of the analysis may be questioned. The results for the R lists appear clear from the plot in Fig. 1. Therefore, a separate analysis was performed on the scores for the four W lists. The F (3,76) = 8.84, $p < .01$, indicates that it is very unlikely that these four means came from the same population. The nonmonotonic nature of the function is indicated by the fact that the means for Lists 2W and 6W differed significantly ($t = 3.48$), as did those for Lists 6W and 12W ($t = 2.80$).

DISCUSSION

A commonsense approach to an interpretation of the nearly perfect performance for the *Ss* given List 2R could be that *S* learns on the study trial that there are only two correct words and instructs himself to respond *always* with them. There is no evidence to deny this possibility. However, to be consistent, such

an interpretation must also be applied to the *S*s given List 2W: it should be equally plausible for *S* to instruct himself *never* to respond with the two W words. Yet, performance on List 2W was appreciably poorer than that on List 2R. On the other hand, the frequency difference for R and W words was less for List 2W than List 2R, which, according to the theory, *would* make List 2W a more difficult task than List 2R.

The results for the R lists were quite in accord with the expectation, from the theory, that as number of R words increases the frequency differential between R and W Words decreases, thus making the discrimination more and more difficult. The results for the W lists also conformed in general to expectations from the theory, but one detail did not. The theory predicts that as number of W words increases, difficulty in learning first increases and then decreases, and this was found. However, the point of maximum difficulty, according to the theory, should have been for List 6W, where the frequency units should be equivalent for the R and W words. The results showed that List 4W was slightly more difficult than 6W. How seriously this failure of the theory should be taken is not clear. It is possible, perhaps, that when a frequency differential is initially indiscriminate (as it might have been for both lists), *S* turns to other characteristics or attributes of the words to establish the discrimination.

Summary

The frequency theory of verbal-discrimination learning makes two predictions concerning learning difficulty as number of correct words and number of incorrect words are varied, holding total number of pairs in the lists constant. First, as number of correct words increases, list difficulty should increase correspondingly. Second, as number of incorrect words increases, difficulty should first increase and then decrease. Eight lists were used, each consisting of 12 pairs. In four lists the number of correct words was 2, 4, 6, or 12, with 12 incorrect words for each list. In four other lists the number of incorrect words was 2, 4, 6, or 12, with 12 correct words for each list. Both predictions were confirmed, but the theory failed to account for the point of maximum difficulty as number of incorrect words was varied.

JUDGMENTS BASED ON DIFFERENT FUNCTIONAL RELATIONSHIPS BETWEEN INTERACTING CUES AND A CRITERION

By STANLEY A. SUMMERS, San Fernando Valley State College, RITA C. SUMMERS, IBM Corporation, Los Angeles Scientific Center, and VIRGINIA T. KARKAU, San Fernando Valley State College

Human adaptation is based in large part on the capacity to discover fixed relations among the varying conditions of the environment, as a means to interpret these conditions accurately and to deal with them appropriately. Certain conditions are useful cues because of their relationship to other, more intrinsically important conditions. Relationships that may be described by mathematical functions link such cues to those conditions that individuals need to predict or interpret. In most situations, apparently, judgments are based on several cues, and the joint function relating the criterion to these cues may be quite complex.

In the experiment reported here, the form of a functional relationship between cues and criterion was varied. Different patterns of criterion values were presented as feedback to different groups of *Ss*. In this way different functions were established between the two stimulus dimensions which served as cues and the criterion value which was to be predicted by the *Ss*. For one group the correct value of the predicted variable depended on the sum of the two cue values; for another group, on the product; for the third and fourth groups one cue value was raised to a power dependent on the other cue value. Each of these particular functions—additive, multiplicative, and exponential—represents a distinct way in which two cues may interact. The experiment asked whether learning would occur under each of these forms of relationship between cues and criterion and whether the accuracy of response would differ. It also asked whether the *Ss* would use curvilinear aspects of the relationships and whether the relation of their learned responses to the cues would tend to match the relationship presented rather than to have some fixed form, such as linear.

* Received for publication October 21, 1968. The research reported was supported in part by Grant M-6410 (A) from the National Institute of Mental Health.

The problem of judgments based on interacting cues was considered by Thorndike.[1] He emphasized complex relationships and cue interactions in the judgment of job fitness and praised, over the existing systems of formal evaluation, the human judge's ability to utilize complex relationships: "The formal systems of the past have used symptoms merely additively. . . . They . . . have not sensed the importance of the multiplying effect of certain traits upon others. The competent impressionistic judge of men does respond to these interrelations of the facts and sums up in his estimate a consideration of each in the light of the others." Several studies using Brunswik's lens model,[2] which provides a correlational framework to measure relations between cues and predicted variable, have demonstrated successful utilization of multiple cues, whether each cue was related in a linear or nonlinear way to the predicted variable. In another study, this one involving the interaction of cues, *S*s learned to estimate accurately the worth of ellipses that varied in size, shape, and color.[3] In the present experiment, cue values were embodied in naturalistic stimulus materials; the cues were characteristics of projected color slides designed to represent highly magnified samples of blood. Different relationships were established between these characteristics of a single series of slides and the varied quantitative information fed back to different groups of *S*s.

METHOD

Task and subjects. The *S* was required to estimate the age in hours of the simulated blood cells shown in each of 640 color slides. After making each response he was given feedback as to the 'correct' age. Each of the four groups of *S*s was presented with the same series of projected slides, but each group was given a different set of feedback values. The *S*s were 64 student volunteers from undergraduate classes.

Materials. Each of the 640 stimulus slides was identified by a letter combination (AA, AB, . . .) which corresponded to a label on *S*'s answer sheet. The slides had been photographed in a sequence that was random with respect to cue values. Each slide contained two relevant cues, each of which had eight possible values. Cue 1 was based on the amount of crushed dark glass scattered across the background. Cue 2 was the size of the rough gray disks that appeared in each slide. There were also nonrelevant variations in exact hue of

[1] E. L. Thorndike, Fundamental theorems in judging men, *J. appl. Psychol.*, 2, 1918, 75.

[2] Egon Brunswik, *Perception and the Representative Design of Psychological Experiments*, 1956.

[3] D. B. Yntema and W. S. Torgerson, Man-computer cooperation in decisions requiring common sense, *IRE Trans. Hum. Factors Electron.*, 1961, HFE-2, 20–26.

background, mottling of background, number of disks, number of small red marbles shown in the slide, and positioning of all the elements. Each of the 640 slides was physically unique, although each possible combination of the cue values appeared 10 times (once in each block of 64 trials).

Ss wrote their responses on cardboard strips which were inserted along with feedback sheets into 8½ × 11 in. transparent plastic envelopes. These feedback sheets were arranged so that after writing an answer on the cardboard strip, *S* could move the strip and then see the correct answer for that trial through a hole in the strip. This method provided immediate feedback after every response and also produced a record of every response. Each *S* used 10 plastic envelopes, each envelope containing the answers for one block of trials.

Procedure. On each trial a slide was projected for 15 sec. Each *S* received 10 blocks of 64 trials, five blocks on each of two consecutive days. Each block of 64 slides contained all possible combinations of the eight values of Cue 1 and the eight values of Cue 2.

Instructions began:

This is an experiment on learning. Today and tomorrow you will be learning a judgment task. You will see a series—a long series—of color slides showing simulated, highly magnified blood cells. Your task will be to estimate the age of the cells shown in each slide. After you make your estimation you will be shown the correct age. The task is difficult, and you probably will need feedback from many slides before you learn to make good estimations. It is extremely difficult to get exactly the right answer for each slide, but the closer you come the better. Your goal should be to come as close as you can.

The slides you will see show blood cells anywhere from 0 to 100 hours old. Your task is to learn to estimate age (between 0 and 100 hours), using as cues some of the features that vary from slide to slide.

Ss then received detailed instructions on use of the materials. On Blocks 1 and 10, *Ss* responded but received no feedback. On each trial of Blocks 2–9, *S* wrote his estimate of blood-cell age on the cardboard strip. He then pulled back the strip, revealing the correct answer for that trial.

There were four experimental conditions, corresponding to the four functional relationships, and each *S* was assigned to one condition. The *Ss* of different experimental conditions participated together, and all *Ss* received the same treatment, except for the answers presented to them as 'correct' blood-cell age. These answers were computed as follows:

$$\text{Condition 1, Age} = \frac{100}{14}(c_1 + c_2 - 2) + v;$$

$$\text{Condition 2, Age} = \frac{100}{63}(c_1c_2 - 1) + v;$$

$$\text{Condition 3, Age} = \frac{100}{7}(c_1^{c_2/8} - 1) + v; \text{ and}$$

$$\text{Condition 4, Age} = \frac{100}{7}(c_2^{c_1/8} - 1) + v;$$

where c_1 and c_2 are the values {1, 2, . . . 8} of Cue 1 and Cue 2, and where v is a number chosen randomly, with equal probability, from the values {−2, −1, 0, 0, 1, 2}. Random variation (v) was added to answers of 13 or greater in order to increase the number of correct answers for each function, but no answer

was allowed to exceed 100. The *S*s in Conditions 3 and 4 were given the same exponential relationship, but with the roles of Cue 1 and Cue 2 reversed. All functions contained a substantial linear component; the correlations of correct value (before addition of v) with Cue 1, with Cue 2, and with both cues *linearly* combined were: Condition 1, .707, .707, 1.0; Condition 2, .665, .665, .941; Condition 3, .574, .666, .879; and Condition 4, .666, .574, .879.

From five to nine *S*s participated at one time, each session involving approximately the same number assigned to each experimental condition. Materials were distributed through the experimental room in a counterbalanced arrangement before *S*s arrived; *S*s were then allowed to choose their own seats. The entire procedure took approximately 1½ hr. on each of the two consecutive days. A 10-min. rest period followed the third experimental block of each day.

Results

The results were analyzed to provide answers for these questions: (*a*) Did *S*s within each condition improve in performance, and was this improvement greater in some conditions than in others? (*b*) Did *S*s make use of the curvilinear as well as the linear aspects of the functions presented to them? (*c*) Was there an initial tendency to respond to the cues in a linear way or some other preferred way? (*d*) Did *S*s eventually match the function to which they had been exposed?

Two measures of response accuracy were computed for each *S*'s responses on each block. The first measure, called the achievement correlation, or r_a, is the Pearson correlation coefficient, computed over the 64 trials within each block, between *S*'s responses and the correct response values for his condition. The second measure, called *MSD*, is the mean square difference, within a block, of *S*'s responses from the correct values. These two measures indicate somewhat different characteristics of response. While *MSD* is a more direct measure of response accuracy, it depends on matching the distribution of correct values for a function as well as on the relation of responses to cues; r_a shows the common variations in cues and responses.

Improvement in performance. Learning curves for all conditions, using r_a, are presented in Fig. 1. These curves show averages of r_a across *S*s, computed using a z transformation. The corresponding curves, using *MSD*, are presented in Fig. 2.

An analysis of variance was performed on the z transformations of the r_a scores for the eight blocks with feedback. The analysis showed a significant overall block effect, $F\ (7{,}420) = 37.92$, $p < .001$, and significant block effects within all conditions, $df =$

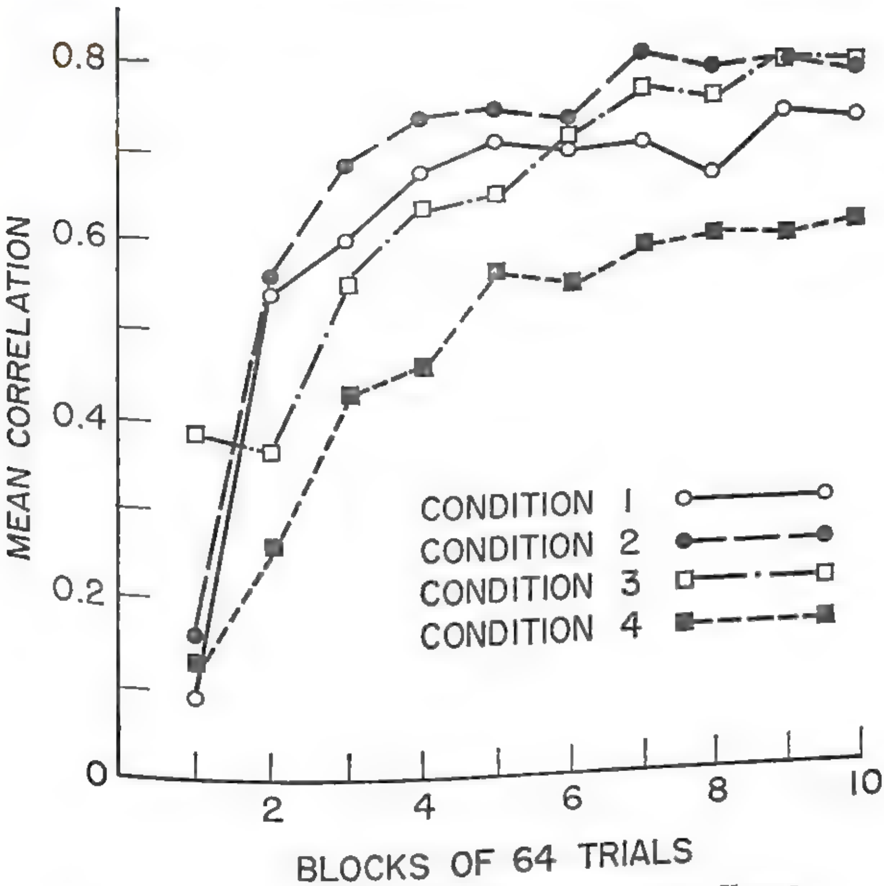


FIG. 1. MEAN CORRELATIONS OF RESPONSES WITH CORRECT VALUES

7,315 for each F: Condition 1, $F = 5.02$, $p < .001$; Condition 2, $F = 9.40$; $p < .001$; Condition 3, $F = 19.92$, $p < .001$; and Condition 4, $F = 8.34$, $p < .001$. The condition effect was also significant, $F\ (3,60) = 6.05$, $p < .001$. Fig. 1 shows that *S*s in Conditions 2 and 3 performed best throughout, and those in Condition 4 performed worst. The Block × Condition effect was significant, $F\ (21,420) = 1.59$, $p < .05$.

Similar results were obtained when an analysis of variance was performed on the *MSD* scores. The overall block effect was significant, $F\ (7,420) = 35.80$, $p < .001$, as were the block effects within all four conditions, $df = 7{,}315$ for each F: Condition 1, $F = 3.54$, $p < .01$; Condition 2, $F = 5.61$, $p < .001$; Condition 3, $F = 15.49$, $p < .001$; and Condition 4, $F = 16.80$, $p < .001$. The condition effect was significant, $F\ (3,60) = 3.34$, $p < .05$, and the

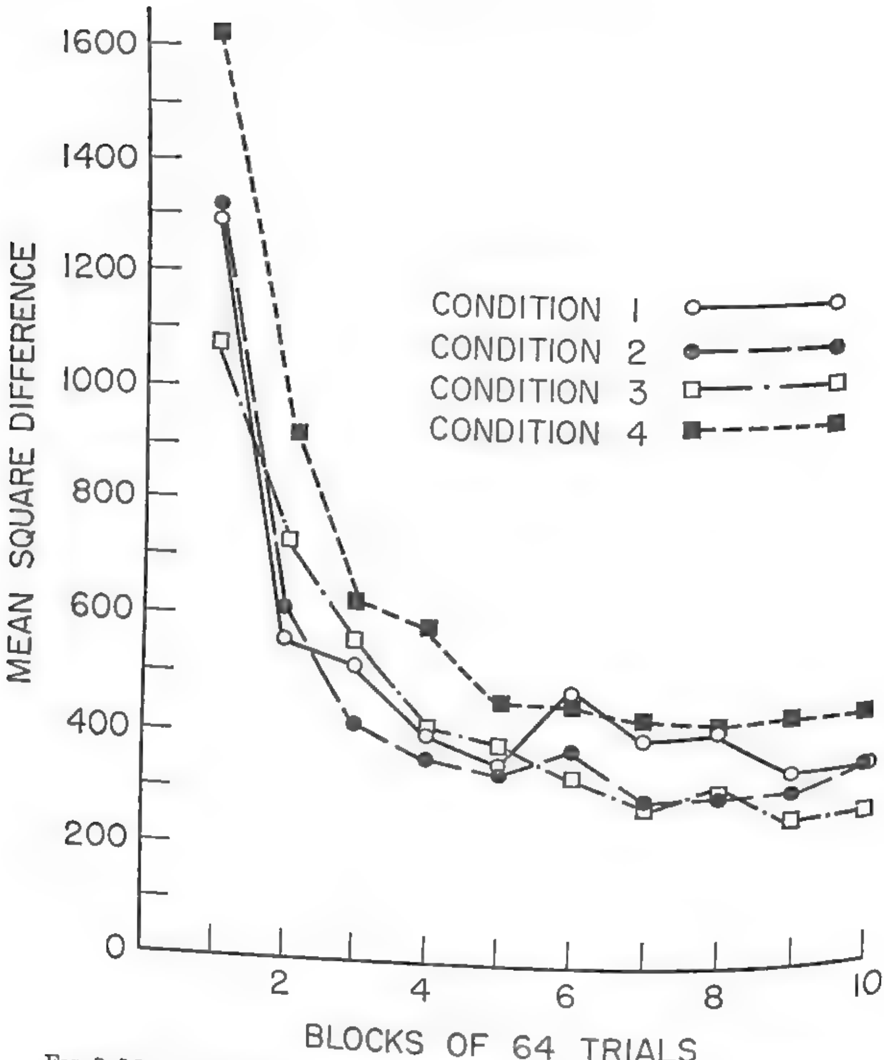


FIG. 2. MEAN SQUARE DIFFERENCES OF RESPONSES FROM CORRECT VALUES

Block × Condition effect was also significant, F (21,420) = 1.88, $p < .05$. Both r_a and *MSD* analyses thus showed that performance improved over blocks, for all types of feedback, that response accuracy differed across conditions, and that there was differential learning among the conditions.

Use of curvilinear aspects. Hursch, Hammond, and Hursch described a technique for determining whether *S*s are making correct use of the nonlinear relations between cues and the variable to be

predicted.[4] The measure C is defined as the correlation between two sets of residual variances, one in the predicted variable and the other in S's responses. These residuals remain after removal of the variance accounted for by multiple correlation with the cues. A high C value thus indicates that S is correctly using a nonlinear relation between the cues and the correct response values.

If the environmental relationships are completely known, C can be computed from response data using the formula:[5]

$$r_a = \frac{R^2_e + R^2_s - d^2_r}{2} + C\sqrt{(1 - R^2_e)(1 - R^2_s)}$$

where r_a is the achievement correlation, R^2_e is the multiple correlation between cues and correct value, computed for each experimental condition, R^2_s is the multiple correlation between cues and S's responses, computed for each block, and d^2_r is the sum of the squared differences between utilization coefficients (correlations of each cue with response) and validity coefficients (correlations of each cue with correct value), computed for each block.

Accordingly, a C value was computed for each block of responses of each S. Fig. 3 shows average C values for each condition (computed using a z transformation). An analysis of variance was performed on the C values of Conditions 2, 3, and 4 (Condition 1 was omitted because it was linear, and all C values were exactly zero). The block effect was significant overall, F (7,315) $= 18.37$, $p < .001$, and within conditions, $df = 7{,}315$ for each F: Condition 2, $F = 4.29$, $p < .001$; Condition 3, $F = 14.13$, $p < .001$; and Condition 4, $F = 3.57$, $p < .01$. The condition effect was significant, F (2,45) $= 7.65$, $p < .01$, as was the Block $\times$ Condition effect, F (14,315) $= 1.82$, $p < .05$. The results show that Ss in all the nonlinear conditions made increasingly correct use of nonlinear relationships.

Initial tendency. Each S's responses on Block 1 (before feedback) were correlated with *all four* sets of correct values. An analysis of variance of z transformations of these correlations showed that responses correlated differently with the various functions, F (3,189) $= 7.01$, $p < .001$. The mean correlations were: Function 1, .24; Function 2, .21; Function 3, .21; and Function 4, .19. Of

[4] C. J. Hursch, K. R. Hammond, and J. L. Hursch, Some methodological considerations in multiple-cue probability studies, *Psychol. Rev.*, 71, 1964, 42–60.

[5] Hursch, Hammond, and Hursch, *op. cit.*, 52. See Formula 23.

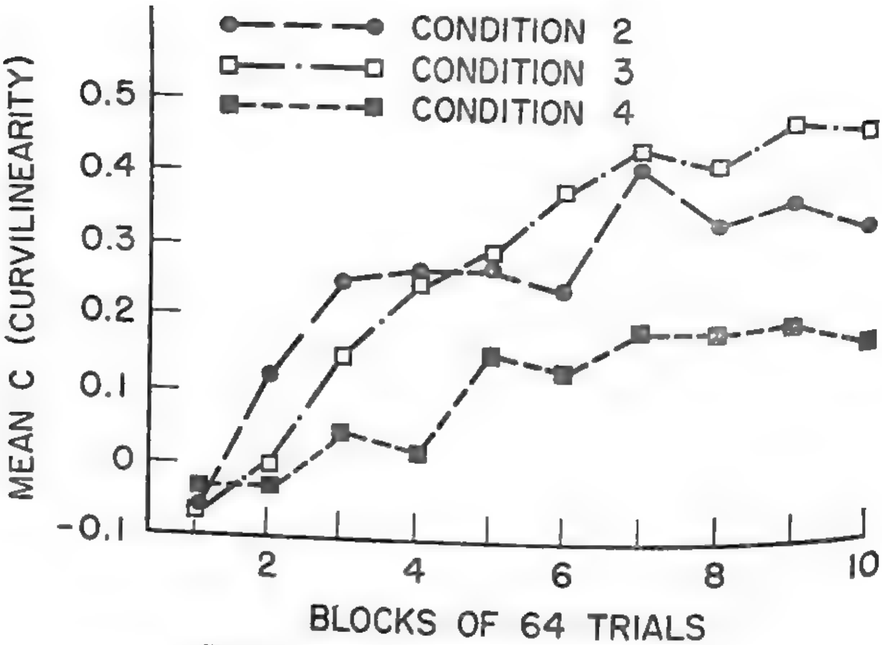


FIG. 3. MEAN *C* VALUES (CURVILINEARITY)

the 64 *S*s, 44 had their smallest *MSD* scores for the linear function. Responses on Block 1, then, were closest to the linear function.

Function matching. On Block 9, the last block with feedback, the responses of 37 *S*s correlated more highly with the particular function presented to them than with any of the other three functions. The responses of 51 *S*s had the smallest *MSD* scores for their own function. Chi-square tests were performed for departures within each condition from the number of *S*s expected by chance to have the closest match (highest r_a or lowest *MSD*) with their function. For r_a, significant χ^2's were obtained for all but Condition 4: Condition 1, χ^2 (1) = 36.8, $p < .001$; Condition 2, χ^2 (1) = 4.1, $p < .05$; and Condition 3, χ^2 (1) = 14.1, $p < .001$. For *MSD*, significant χ^2's were obtained for all four conditions: Condition 1, χ^2 (1) = 44.1, $p < .001$; Condition 2, χ^2 (1) = 14.1, $p < .001$; Condition 3, χ^2 (1) = 30.1, $p < .001$; and Condition 4, χ^2 (1) = 10.1, $p < .01$. In spite of an initial tendency to use a linear function, the *S*s generally came to match their own function most closely.

DISCUSSION

In each condition the *S*s learned to use successfully the particular relationships presented to them. Also, they learned to use the non-

linear aspects of these relationships, responding to cue interactions. The *Ss* used nonlinear aspects in spite of the fine discrimination required; a large linear component was present in all functions, allowing quite accurate predictions from purely linear responses. The *Ss* clearly were not limited to a linear response mode. Although they used each functional relationship successfully, the *Ss* did not learn to make completely correct responses. The interaction of physical cue characteristics with a functional relationship may be as important as the form of the relationship itself in determining accuracy. Although Functions 3 and 4 had the same form, the role of the two cues was reversed; responses in the last two blocks were most accurate for Function 3 and least accurate for Function 4. In spite of such limitations, however, the overall accuracy of the *Ss* was impressive.

This experiment was characterized by the perceptual difficulty and the complexity of the task, and by the large number of trials. The results therefore suggest that humans in nonexperimental settings also learn to use whatever valid relationships are to be found. Such flexibility would count much more in successful adaptation to an environment than would any fixed mode of response.

Summary

The form of the function relating cue and criterion values was varied. The cues were two stimulus dimensions on slides of simulated blood cells; the criterion was the age of the cells. Each *S* in four groups of 16 undergraduates was presented with one particular relationship (linear, multiplicative, or exponential) and attempted to predict the criterion values from the cues. The *Ss* learned over the 640 trials to use the relationship to which they were exposed, as indicated by increased correlations between responses and criterion and by decreased differences of response from criterion values. The valid use of nonlinear components increased over trials and, in spite of an initial tendency to use a linear function, the *Ss* generally came to match their own function most closely.

INTRA- AND EXTRADIMENSIONAL SHIFTS IN RETARDATES AS A FUNCTION OF DIMENSIONAL PREFERENCE

By JOSEPH C. CAMPIONE, University of Washington

In the simplest form of concept-shift experiment, the *S* learns a series of two problems, an original problem and a transfer problem. The type of shift is defined by the relation between the relevant cues of the original and transfer problems. In an intradimensional (ID) shift, the relevant cues of the two problems lie along the same dimension. In an extradimensional (ED) shift, the relevant cues of the transfer problem lie along a different dimension than those of the original problem; *e.g.* the positive cue is changed from 'red' to 'square.' A large number of experimenters have compared performance on ID and ED shifts and have obtained consistent results. ID shifts are learned more quickly than ED shifts by rats,[1] monkeys,[2] preschool children,[3] retarded children,[4] grade-school children,[5] and adults.[6] These data are consistent with chaining theories of discriminative learning in which the first link of the chain is a response to some dimension of the discriminanda.[7] The second

* Received for publication October 25, 1968. The research reported was supported by Grant M-1099 from the National Institute of Mental Health of the U. S. Public Health Service. The author is indebted to Drs. David Zeaman and Betty House for their suggestions and aid throughout the conduct of the research. Grateful acknowledgment is also made for the cooperation of Francis P. Kelly, Superintendent of the Mansfield State Training School, and Louis Boly, Director of Training of the Longley School.

[1] B. E. Shepp and P. D. Eimas, Intradimensional and extradimensional shifts in the rat, *J. comp. physiol. Psychol.*, 57, 1964, 357–361.

[2] B. E. Shepp and A. M. Schrier, Performance of macaque monkeys on successive intradimensional and extradimensional shifts, *J. comp. physiol. Psychol.*, 1968, in press.

[3] D. J. Dickerson, Performance of preschool children on three discrimination shifts, *Psychon. Sci.*, 4, 1966, 417–418.

[4] J. C. Campione, L. M. Hyman, and David Zeaman, Dimensional shifts and reversals in retardate discrimination learning, *J. exp. child Psychol.*, 2, 1965, 255–263.

[5] P. D. Eimas, Effects of overtraining and age on intradimensional and extradimensional shifts, *J. exp. child Psychol.*, 3, 1966, 348–355.

[6] K. H. Kurtz, Discrimination of complex stimuli: The relationship of training to test stimuli, *J. exp. Psychol.*, 50, 1955, 283–292.

[7] N. S. Sutherland, Stimulus analyzing mechanisms, in *Proceedings of a Symposium on the Mechanization of Thought Processes*, Vol. 2, 1959, 575–609; David Zeaman and B. J. House, The role of attention in retardate discrimination

link (dubbed "instrumental learning" by Zeaman and House) consists of the attachment of responses to the cues (*e.g.* red and green) of the observed dimension (*e.g.* color).

The present experiment was an attempt to separate two classes of chaining theory which differ in their choice of a theoretical dependent variable. Theorists have assumed that reward and nonreward affect either the probability of attending to a dimension or the strength of a dimension. While it may appear that the choice depends primarily upon mathematical considerations (probabilities being easier to work with and easier to relate to behavior), such is not entirely the case, since the choice can lead to different sets of predictions. The variable of interest, dimensional preference (which has been shown to affect performance on reversal and nonreversal shifts),[8] is relevant because different predictions about its effects on the ID-ED comparison can be derived from different versions of chaining theory. According to the one-look model of Zeaman and House, the effect of rewarding *S* for responding to a particular dimension is to increase *S*'s probability of attending to that dimension on subsequent trials and to decrease his probability of attending to any of the competing dimensions.[9] In this model, the sum of the observing-response probabilities must be unity. For *S* to be consistently choosing the positive stimulus, his probability of attending to the relevant dimension must be high and his probability of attending to any of the competing dimensions very low. Consequently, performance differences between ID and ED shifts should be large and independent of the initial observing-response probabilities and, therefore, independent of the original dimensional preferences of *S*. The only effect of dimensional preference should be found on original learning. Since *S*'s initial probability of attending to a preferred dimension is higher than his probability of attending to a nonpreferred dimension (by definition), he should learn a problem faster when his preferred dimension is relevant.

In the model proposed by Sutherland, the dimensions (which correspond to his "stimulus analysing mechanisms") are assumed

learning, in N. R. Ellis (ed.), *Handbook of Mental Deficiency*, 1963, 159–223; E. P. Lovejoy, An attention theory of discrimination learning, *J. math. Psychol.*, 2, 1965, 342–362.

[8] S. S. Smiley and M. W. Weir, Role of dimensional dominance in reversal and nonreversal shift behavior, *J. exp. child Psychol.*, 4, 1966, 296–307; L. W. Heal, M. L. Bransky, and R. L. Mankinen, The role of dimensional preference in reversal and nonreversal shifts of retardates, *Psychon. Sci.*, 6, 1966, 509–510.

[9] Zeaman and House, *op. cit.*, 169 ff.

to differ in strength.[10] Using a measure of strength rather than of probability allows these strengths to vary independently, an increase in the strength of one dimension not being accompanied by decreases in the strengths of the competing dimensions. The data generally indicate that instrumental learning proceeds rapidly, so that once the relevant dimension becomes dominant, the problem should be quickly solved. Also, once the relevant dimension becomes dominant, *S*'s responses should usually be correct; thus, through reinforcement, the strength of the dimension should increase further. It is assumed in the present paper that the increment in the strength of the relevant dimension from the time it becomes dominant until the problem is solved is some constant x. Further, while Sutherland did not propose any specific rule relating strengths to behavior, it is assumed here that for dimension B weaker than dimension A, the number of trials required for B to become dominant over A is related to the difference between the strengths of dimensions A and B.

Assume we have two dimensions, A (with original strength s_a) and B (with original strength s_b). Assume further that $s_a > s_b$ and that $s_a - s_b = d$. If we train *S* on an original problem in which dimension A is relevant, the problem should be solved quickly, since dimension A is dominant from the outset of the problem. Once the problem is solved, the strength of dimension A has increased by an amount x. Thus, at the end of the original problem, the difference between the strengths of dimensions A and B should be $d + x$. This difference should be a measure of the differential rates at which ID and ED shifts are solved following training on a problem in which dimension A is relevant. An ID shift should be solved rapidly, since the relevant dimension (A) is dominant at the outset of the problem. In the case of an ED shift, dimension B must become dominant over dimension A before the problem can be solved. The number of trials required for B to become dominant should be directly related to the difference between the strengths of dimensions A and B ($d + x$) when the shift problem is introduced.

Consider now the case in which we originally train *S* on a problem in which dimension B is relevant. This problem should be solved slowly, since dimension B must become dominant before the problem can be solved. As soon as B becomes dominant, the problem should be solved quickly, during which time the strength of dimension

[10] Sutherland, *loc. cit.*

B should increase by x. Thus, the difference between the strengths of dimensions A and B following solution of the original problem should be approximately x. This difference should be a measure of the differential rates at which ID and ED shifts are solved following training on a problem in which dimension B is relevant. Since $d + x > x$, we should expect a larger difference between ID and ED shifts after training on dimension A (the originally preferred dimension) than after training on dimension B. The magnitude of this effect should depend on d, a measure of the original preference. Finally, a gross estimate of d can be obtained by considering the original learning data. To the extent that d is large, *S*s should solve the original problem more quickly when dimension A is relevant than when dimension B is relevant.

Rather than assessing dimensional preferences for each *S*, an attempt was made to manipulate dimensional preference experimentally. Shepp and Zeaman have shown that the probability of attending to any dimension depends in part on the physical difference between the cues along that dimension.[11] In the present experiment, the stimuli differed along the dimensions of color, form, and size, with the differences in size being fairly large. In this way, it was assumed that size would be the dominant dimension for most *S*s. Half the *S*s were originally trained with size as the relevant dimension, the remainder with form as the relevant dimension. These two groups were then subdivided, with half the *S*s in each group getting an ID shift, and the remainder getting an ED shift. Thus, the design of the experiment was a 2 (Dimension: size vs. form relevant on the original problem) × 2 (Shift: ID vs. ED) factorial. A test of the competing theories could thus be made by considering the Shift × Dimension interaction. The probabilistic theory proposed by Zeaman and House does not predict a significant interaction effect, whereas the strength model outlined above would predict the interaction.

Method

Subjects. The *S*s were mentally defective children attending the Longley School of the Mansfield State Training School. A total of 32 children, 20 boys and 12 girls, completed the experiment. Their MAs ranged from 40 to 65 months, with a mean of 50; their IQs from 30 to 67, with a mean of 48. The *S*s were divided into four groups, matched with respect to MA and IQ.

Apparatus and stimuli. A modified version of the Wisconsin General Test

[11] B. E. Shepp and David Zeaman, Discrimination learning of size and brightness by retardates, *J. comp. physiol. Psychol.*, 62, 1966, 55–59.

Apparatus was used. The *S* and *E* sat facing each other with a one-way vision screen interposed. A 30 × 12 in. stimulus tray containing two circular food wells 2 in. in diameter could be directly presented in front of *S* by pushing it through a space beneath the one-way screen. The entire apparatus, excepting the one-way screen, was painted gray. Stimuli were two-dimensional patterns differing in color, form, and size. Small stimuli were cut from a square 1 in. on a side, large stimuli from a square 1½ in. on a side. The forms used were regular geometric patterns (*e.g.* circle, square, cross); the colors used were red, green, blue, yellow, violet, and black.

Pretraining. Each S was brought individually to the experimental room where he was told he was going to play the "candy game." Pretraining was begun immediately, the reward for a correct response being an M & M candy. In addition to the candy, correct responses were verbally rewarded by "good" and incorrect responses 'punished' with "no." Each *S* was run on a two-choice visual discrimination employing two randomly selected 'junk' stimuli (stimuli differing multidimensionally in size, shape, color, texture, etc.). The position of the correct stimulus was varied according to a Gellermann series. A correction procedure was used throughout the experiment. Training was continued for 25 trials a day until *S* reached a criterion of 20 or more correct responses on a given day. *Ss* not reaching criterion in four days (100 trials) were dropped from the experiment. A total of four *Ss* was dropped for this reason.

Original learning. Half the *Ss* were given a two-choice visual discrimination problem in which size (form) was relevant, with form and color (size and color) irrelevant and variable within trials. With three visual dimensions and position varying, there was a total of eight possible stimulus settings. Each of these settings appeared twice in each block of 16 trials. The order within a block was random except that no setting appeared on two consecutive trials and that there were never more than three consecutive rights or lefts. The colors and forms were selected randomly for each *S*. Of those *Ss* who were originally trained on a size problem, half were rewarded for choice of the large stimulus, and half for choice of the small stimulus. Training was continued to a criterion of 10 consecutive correct responses. *Ss* not reaching criterion in four days (100 trials) were dropped from the experiment. A total of eight *Ss* failed to reach criterion on the original problem, seven failing a form problem and one a size problem.

Shift problem. Immediately after *S* reached criterion on the original problem, the shift problem was introduced. Again, one dimension was relevant, with the others irrelevant and variable within trials. Entirely new color and form cues were introduced, but the size differences were the same as in the original problem. (If relative size serves as one dimension, it is impossible to introduce entirely new cues.) Further, it was deemed important to keep the physical difference between the sizes the same on the original and shift problems. Of the *Ss* who had had form relevant on the original problem, half were given a shift problem in which form was relevant (ID shift), the remaining half a problem in which size was relevant (ED shift). For the ID group, the positive stimulus was determined randomly for each *S*; for the ED group, half were rewarded for choosing the large stimulus, the remaining half for the small stimulus. Similarly, of the *Ss* who had had size relevant on the original problem, half were given an ID shift and half an ED shift. The ID group consisted of

four *Ss* who had been rewarded for choice of the large stimulus on the original problem and four who had been rewarded for choice of the small stimulus. Half these *Ss* were rewarded for making the same choice on the shift problem, the remaining half for choosing the other size. In essence, half the *Ss* were given a continuation of the original problem (with novel irrelevant cues), and half were given a reversal (also with novel irrelevant cues). For the *Ss* in the ED group, the positive stimulus was determined randomly for each *S*.

Results

To correct for heterogeneity of variance, all analyses were carried out using logarithmic transformations of the raw data. The dependent variable was log (trials to criterion + 1).

Original learning. A 2 (Dimension) × 2 (Shift) factorial analysis of variance was performed on the transformed original-learning data. The only significant effect was due to dimension ($F = 10.63$, $df = 1/28$, $p < .001$), indicating faster learning when size was the relevant dimension. The *Fs* associated with shift and the Shift × Dimension interaction were less than one.

Shift problem. The geometric means (trials to criterion + 1) for the four groups were as follows: after training on a size problem, the ID shift required 5.4 trials, the ED shift 34.1 trials; after training on a form problem, the ID shift was learned in 5.7 trials, the ED shift in 7.3 trials. Again, a 2 × 2 factorial analysis of variance was performed on the transformed data. The main effects of dimension and shift were reliable ($F = 9.64$, $df = 1/28$, $p < .001$ and $F = 18.16$, $df = 1/28$, $p < .001$, respectively), as was the Shift × Dimension interaction ($F = 10.37$, $df = 1/28$, $p < .001$). To follow up the interaction, *t* tests (based on the analysis of variance error term) were carried out comparing ID and ED shift performance following training on either a size or form problem. Following training on a size problem, ID shifts were learned significantly more quickly than ED shifts ($t = 5.33$, $df = 28$, $p < .001$), whereas the shifts did not differ in difficulty following training on a form problem ($t < 1$).

Discussion

The results of the original-learning analysis indicate that the manipulation of dimensional preference was successful. *Ss* learned the size problem significantly more quickly than the form problem, indicating that size was the preferred dimension for most of the *Ss*. A similar conclusion is reached if we consider the number of

failures on the original problem. Of the eight *Ss* who failed to reach criterion on the original problem, seven had form as the relevant dimension. Using a binomial test, this outcome has a probability of occurring of less than .05 (one-tailed test).

Considering the shift data, the reliable Shift × Dimension interaction is consistent with the strength model outlined above, but at odds with the one-look probability model proposed by Zeaman and House. These results, then, support a model in which the acquisition and extinction of selective attentional responses proceed independently, increases in the strength of one dimension not being accompanied by decreases in the strengths of the competing dimensions. Before this interpretation can be accepted, however, several alternative explanations should also be considered. While the results are inconsistent with the one-look model of Zeaman and House, they may be compatible with a multiple-look model. If, for example, the result of training *S* on a nonpreferred dimension is to have him attend to both the preferred and nonpreferred dimensions on subsequent problems, the outcome of the present experiment could be predicted.

One further explanation also seems plausible. We might assume that the strengths of the dimensions (or probabilities of attending to the dimensions) do not transfer perfectly to new problems or, more specifically, that the introduction of a new problem leads to a change in the strengths of the competing dimensions—a change in the direction of returning the strengths to their original values. There are some data which suggest that this type of effect does occur, at least with instrumental responses. (It is generally assumed by chaining theorists that the laws which govern instrumental responses also apply to attentional responses.) For example, in an earlier series of experiments, the author was led to the conclusion that *Ss'* responses in a discrimination task were in part contingent upon the specific stimuli presented.[12] In learning a discrimination involving stimuli A and B (A positive), *Ss* would learn, at least in part, to "approach A, when A and B are presented." The strength of the tendency to approach A changed when A was presented with some other stimulus. Further, Mowrer and Jones invoked a similar principle in trying to account for the effects of partial reinforcement and patterned reinforcement on resistance to extinction.[13]

[12] J. C. Campione, Transitivity and choice behavior, *J. exp. child Psychol.* 1968, in press.

[13] O. H. Mowrer and H. M. Jones, Habit strength as a function of the pattern of reinforcement, *J. exp. Psychol.*, 35, 1945, 293–311.

A theory which includes a similar feature has recently been proposed by Lovejoy.[14] He postulated that the strength of a dimension is determined jointly by three components: the dimension's distinctiveness (which remains constant throughout the experiment), its directable distinctiveness (which varies as a function of reward and nonreward within a problem), and the extent to which *S* has formed "opinions" about the cues of the dimension by associating them with different outcomes. The first component can include the original preferences of *S*. Further, since Lovejoy allows the directable distinctiveness to change only a slight amount, the major fact which increases attention to the relevant dimension within a problem stems from *S*'s changing opinions about the cues along that dimension. When a new problem is introduced, the strength of the previously relevant dimension must decrease sharply, since *S* can have no opinions about the novel cues. Thus, with the introduction of a new problem, attention is determined entirely by the distinctiveness and directable distinctiveness of the dimensions, allowing the original preferences to influence shift behavior, as they did in the present experiment.

While Lovejoy's model can account for the outcome of the present experiment quite readily, it may be able to do so only at a cost. The model includes a large number of parameters, and the results of this experiment are likely consistent with only some sets of parameter values. For example, the large effect of preference on shift behavior obtained here suggests that the distinctiveness of the dimensions is large relative to the directable distinctiveness. What is not clear is whether or not selection of these sets of parameter values adversely affects predictions in other situations. For example, can Lovejoy use these parameter values to account for the continued improvement over successive ID shifts reported by Shepp and Schrier[15] or for the similar effect obtained with successive reversals? It appears to the writer that one-trial reversals or the very fast learning obtained in learning-set experiments is impossible unless the distinctivenesses of the dimensions are allowed to change within an experiment.

Finally, the present results are consistent with those of Smiley and Weir[16] and of Heal, Bransky, and Mankinen[17] in demonstrating that dimensional preferences do affect the transfer behavior of

[14] E. P. Lovejoy, *Attention in Discrimination Learning: A Point of View and a Theory*, 1968.
[15] Shepp and Schrier, *loc. cit.*
[16] Smiley and Weir, *loc. cit.*
[17] Heal, Bransky, and Mankinen, *loc. cit.*

normal and retarded children. Also consistent is the fact that retarded *S*s frequently reverted to position habits even after they solved a number of problems in which position was irrelevant.

Summary

The behavior of retarded children on intradimensional and extradimensional shifts was investigated following training on either a preferred dimension or a nonpreferred dimension. ID shifts were solved reliably faster than ED shifts when the original training was on a preferred dimension, whereas there was no reliable difference when the original training was on a nonpreferred dimension. These results were seen as consistent with a theory in which the acquisition and extinction of selective attentional responses proceed independently. A number of alternative explanations were also proposed, and further research can be directed toward separating these explanations.

A CONTEXTUAL EFFECT IN JUDGMENTS OF VISUAL NUMEROUSNESS

By DONALD GRANBERG, Northern Illinois University, and JOHN ABOUD, JR., Villanova University

Estimating the number of objects in a given situation without actually counting them is a common experience. Reporters and officials may guess the number of people attending a political rally, farmers might estimate grain held in storage, and announcers of athletic events frequently guess the number of people present. The factors that at a given time interact to determine an estimate constitute the frame of reference for that situation.[1] Different factors will be salient in different situations. Reporters, participants, and police officials are all undoubtedly influenced by their attitudes as internal factors in their estimates of the number of people present at a political demonstration. As the actual number becomes more difficult to estimate, the influence of the attitudes becomes more pronounced. The farmer estimating grain in a building may or may not be neutral. External factors, such as the size of the building and the extent to which the building is filled, can also play a role in determining his judgment. If the actual capacity of the building is known, this can increase the accuracy of a judgment. Similarly, observers of athletic events are influenced in their estimates of size of crowd by the actual capacity of the stadium.

Estimates of the number of objects in a given situation, without actually counting the objects, are referred to as judgments of numerousness.[2] Several studies have been reported which document associations between stimulus arrangements and judgments of numerousness. Mokre found that his *S*s decreased their judgments of numerousness as the distance between the dots being judged was increased.[3] Taves found, among other things, that dots ar-

* Received for publication February 10, 1968. Dr. Robert Freeman's comments on an earlier study led to the variations used in the experiment reported. JoAn Garofalo and Nanci August assisted in the preparation of stimulus materials.

[1] M. Sherif and C. Sherif, *An Outline of Social Psychology*, 1956, Ch. 2 and 3.

[2] S. S. Stevens, *Handbook of Experimental Psychology*, 1951, 5, 22, and 28.

[3] H. Mokre, Ueber den Einfluss von Grosse und Abstand der Elemente auf die Mengenauffassung, *Z. Psychol.*, 105, 1927, 195–225.

ranged in a circle (good form) were judged as less than the same number of dots scattered randomly.[4] Bevan, Maier, and Helson varied the size of the context and the figure-ground relationship (through instructions) in an experiment in which *Ss* estimated the number of beans in a jar.[5] They found that the estimates were consistently larger with larger containers. Thus, with the actual number of beans held constant, the less full larger jar was believed to contain more beans than the more full smaller jar. In an experiment involving the judgment of numbers of dots, Bevan and Turner found an interaction effect between the size of the frame and instructions as to figure and ground.[6] The judgments of numerousness were greatest when the large frame was figure and smallest when the small frame was figure. It is evident that the central motif of Gestalt psychology—that stimulus objects do not have absolute stimulating values but must be considered in relation to the context in which they appear—has been the major theoretical perspective influencing these experiments. Few psychologists now doubt the veracity of this gestalt theme, but it remains to specify the ways in which contextual effects operate.

The purpose of the present paper was to report an experiment employing an alternative form of this contextual effect. It concerned judgments of numerousness when the potential 'capacity' of an array was varied. This capacity was defined by adding not-to-be-counted elements to an array of to-be-counted elements. Judgments were made of the number of blackened circles, varying at the same time the number of circles not blacked in. The authors predicted that the circles not blacked in, or unfilled circles, would constitute a background factor and that the judgments of the number of blacked-in, or filled, circles would be made relative to the number which *could* be filled. The hypothesis was that the judgments of numerousness would decrease as the proportion of filled circles to total circles decreased.

Method

Subjects. Forty-eight *Ss*, 46 girls and 2 boys, from sections of Introductory Psychology participated in the experiment. Each received course credit for doing so.

[4] E. Taves, Two mechanisms for the perception of visual numerousness, *Arch. Psychol.*, 1941, No. 265.

[5] W. Bevan, R. Maier, and H. Helson, The influence of context upon the estimation of number, this Journal, 76, 1963, 464–469.

[6] W. Bevan and E. Turner, Assimilation and contrast in the estimation of number, *J. exp. Psychol.*, 67, 1964, 458–462.

Stimulus materials. Forty slides, 10 for each of four experimental conditions, presented a large white circular field, 24 in. in diameter, against a black background. Within the white field were small circles 0.75 in. in diameter. The actual numbers of the black circles were 18, 20, 22, 24, 26, 28, 30, 32, 34, and 36. These black circles were in all cases scattered randomly within the large field.

In addition to the black circles, there were a number of small unfilled circles. For Condition 0X the slides contained only black circles. For Condition 1X there were an equal number of black and unfilled circles. For Condition 2X there were twice as many unfilled circles as black circles, and for Condition 4X there were four times as many. Thus, the slides containing the maximum number of circles to be judged had 36 black circles in Condition 0X, 36 black and 36 unfilled circles in Condition 1X, 36 black and 72 unfilled circles in Condition 2X, and 36 black and 144 unfilled circles in Condition 4X. In Condition 4X, with 36 black and 144 unfilled circles, the large white field was completely filled. The slides were projected by a Kodak Carousel projector in a semi-darkened room.

Procedure. *Ss* were randomly assigned to one of the four experimental conditions. Three *Ss*, on the average, made their judgments at the same time and were seated at a table about 10 ft. from the screen. They were handed a sheet with blanks numbered from 1 to 30, and they were then read these instructions:

We are studying the ability of people to make judgments of numbers with short exposure times. We are going to present a series of slides on the screen in front of you. On each slide will be a number of completely black circles. No slide has more than 40 or less than 14 circles which are blacked in.[7] Make your estimates without counting or trying to count them, as the time allotted to each slide is not sufficient for counting. All you have to do is take a good look at the picture and then make your guess of the number of circles blacked in. Write your estimates on the paper. There will be two blanks between each actual slide so you will have plenty of time. Are there any questions before we begin?

They were then shown the 10 slides appropriate to their experimental condition. Exposure time was 3.0 sec. in all cases, with an interval of 8.0 sec. between slides. The order of exposure was the same in all conditions: 34, 28, 20, 30, 36, 18, 22, 32, 26, and 24. Following the first exposure, the slides were shown a second time with the order reversed, and then a third time in the original order. *Ss* were not told they were seeing the same slides. Each *S* thus made 30 judgments, three of each slide.

RESULTS

Fig. 1 shows the mean judgments for the first presentation of the slides in each of the four conditions. The predicted linear relationship is clearly supported by these data. Condition 0X, with only black circles, resulted in the largest average judgments at every stimulus number. Condition 4X, which contained the greatest number of unfilled circles, or potential capacity, resulted in the smallest average judgments at nine of the 10 stimulus numbers. With only

[7] These range limits were given because in pilot studies, the variance in judgments was sometimes too large to obtain reliable results.

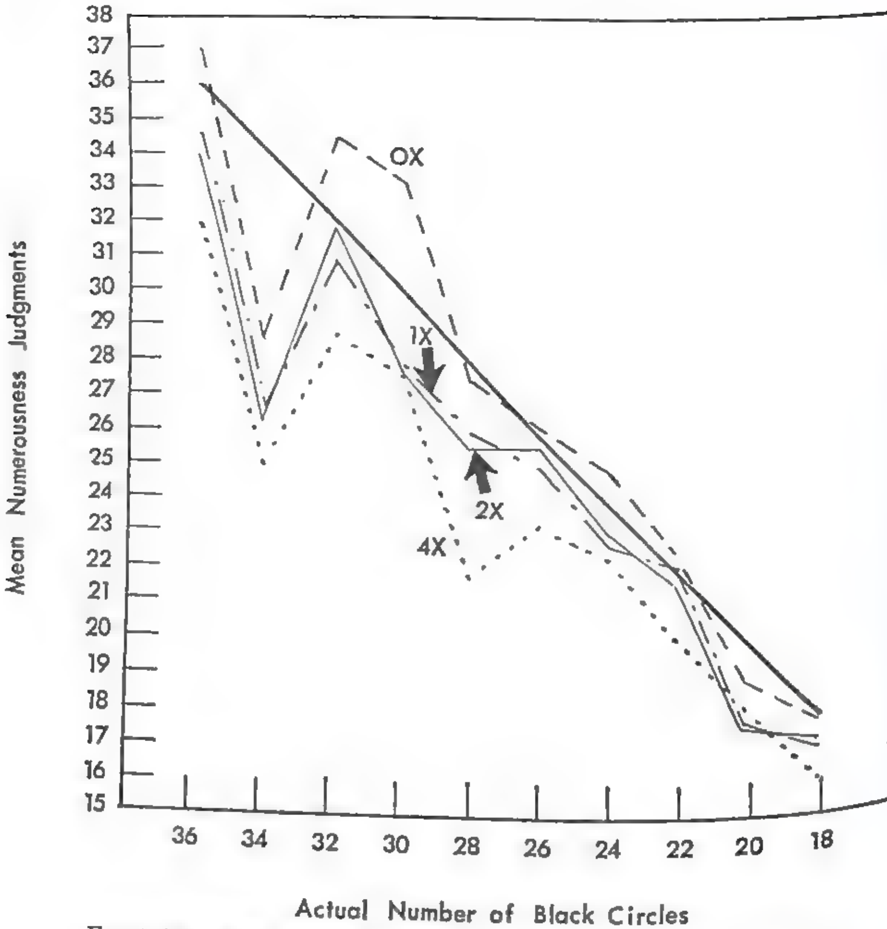


FIG. 1. MEAN JUDGMENTS FOR FIRST PRESENTATION OF ALL SLIDES

one exception, namely at 20, the average judgments for Conditions 1X and 2X fell between the averages for Conditions 0X and 4X. In Condition 0X the average of judgments decreased as the stimulus value decreased, with one exception, that being at 34. This might be accounted for by an order effect since the slide containing 34 was the first one shown.

A score was obtained for each S by adding together his first 10 judgments. An analysis of variance was then computed, comparing the four conditions. An F of 4.24 was obtained which, with 3 and 44 degrees of freedom, was significant at less than the .05 level. The linear component was significant at less than the .01 level and accounted for 90.1% of the between sum of squares.[8]

[8] B. Winer. *Statistical Principles in Experimental Design,* 1962, 73.

On the second and third exposures, the experimental effect weakened considerably. Calculations based upon the total sum of all judgments for each subject resulted in an F of 2.35. Although the results were in the same direction as those in the first 10 judgments, they were not statistically reliable. However, when the median of the three judgments of each S was summed with others in the same condition and the sums of the four conditions ranked, a χ^2 of 12.78 was obtained by a Friedman two-way analysis of variance.[9] This was significant beyond the .01 level with three degrees of freedom. Approximately the same linear relationship lay behind this significance.

The results for the slide containing 36 black circles were especially important for two reasons. First, this is the slide which had the large white field completely filled (in Condition 4X). Second, it is the slide which contained the largest number of circles; hence, it should have been the most difficult to estimate, since it was furthest from what S could count during the exposure time. Fig. 2 shows the mean judgments of the slide containing 36 black circles for the four conditions from the first to the third judgments. Combining the results from the three exposures, an overall F of 7.15 was obtained which, with 3 and 140 degrees of freedom, was significant at less than the .01 level. In addition, as can be seen in Fig. 2, the order of the averages was consistently in line with the hypothesis for each of the three judgments. The Fs for the first and third judgments were 3.35 and 3.12, respectively, each of which was significant beyond the .05 level, the degrees of freedom being 3 and 44 in each case. The results for the second judgment, though in the same direction, were not statistically significant. Fig. 2 shows another relationship which held true for all four conditions. The average judgments increased from the first to the second exposure, then decreased on the third exposure, though not quite back to the level of the first exposure.

Discussion

Fig. 1 showed a dip in the lines which was quite likely due to an order effect. Had it been the purpose of the experiment to study the accuracy of judgments, it definitely would have been necessary to vary the order of presentation, thus controlling for order effects. It may be recalled, however, that the purpose of the present study was limited to demonstrating a contextual effect.

[9] S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*, 1965, 166.

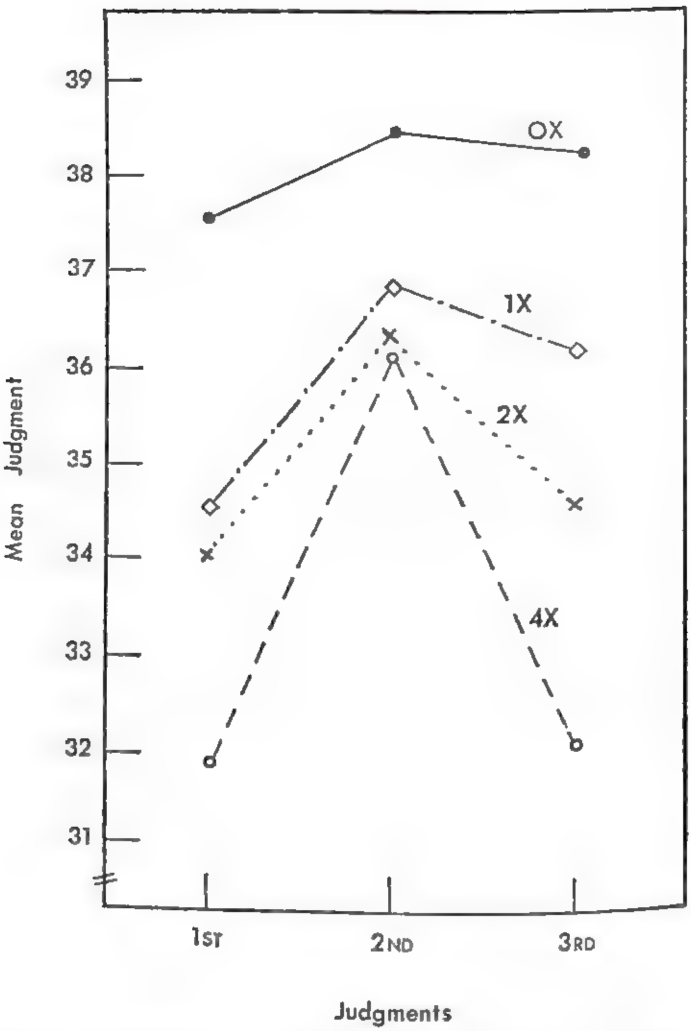


Fig. 2. Mean Judgments for All Presentations of Slide with 36 Black Circles

The results of the present study are consistent with Mokre's earlier study.[10] The two reports document the commonly experienced effect that a number of objects is judged as greater when they are less dispersed or when they fill a larger proportion of the possible space. Both situations are similar to judging the same sized crowd in large and small stadiums. It appears, however, that a different kind of effect was studied by Bevan and his associates. The experiment reported by Bevan *et al.* showed that estimates

[10] Mokre, *loc. cit.*

decreased as the portion of a jar filled increased.[11] The dispersion of the beans in the jar presumably did not vary, the density being high, so their findings are not directly comparable to Mokre's. In addition, their study did not use a jar which was completely filled, and the experiment was complicated further by the figure-ground instructions. With reference to the experimental report by Bevan and Turner, it can be argued that what is important relative to context effects is what occurred in the ground conditions, since in ordinary judgments the frame acts as a ground within which the judgment of the array or figure takes place.[12] It is noteworthy that in the ground conditions, estimates increased as the size of the frame decreased.

[11] Bevan, Maier, and Helson, *loc. cit.*
[12] Bevan and Turner, *loc. cit.*

THE ABSOLUTE AND RELATIVE SIZE CUES TO DISTANCE

By WALTER C. GOGEL,
University of California, Santa Barbara

There is evidence that familiar size as a cue to distance involves not one, but two cue systems. This has been demonstrated by measuring the perceived distance of a familiar object viewed monocularly in an otherwise dark visual field when different retinal sizes of the object are presented successively to *O*. Under these conditions, it is often found that the perceived distance of a particular retinal size differs depending upon whether that size is the first or a later presentation in the series.[1]

The present paper suggests that the first presentation may be considered to involve an *absolute cue* to distance—a cue defined as the size of the familiar object on the eye. The later presentations may be considered to involve both absolute and relative cues to distance—a *relative cue* defined as the difference between the retinal sizes of the familiar objects in the successive presentations. More generally, it may be said that absolute and relative retinal size represent different points on a continuum defined as the time between presentations to *O*. Absolute size weights heavily as a cue in the *first presentation*, which is somewhat temporally isolated from previous visual experience with similar familiar objects. Absolute and relative size as cues to distance combine in *successive presentations*. Relative size weights most heavily in *simultaneous presentation* of two different retinal sizes of the same familiar object. To clarify the proposed differentiation between absolute and relative cues, therefore, the present study examined the effect of temporal separation between presentations on the relation between retinal size and perceived distance by comparing the perceived distances obtained from first, successive, and simultaneous presentations.

* Received for publication August 14, 1967.

[1] W. C. Gogel, B. O. Hartman, and G. S. Harker, The retinal size of a familiar object as a determiner of apparent distance, *Psychol. Monogr.*, 71, 1957 (13, Whole No. 442), 13–14; W. C. Gogel and H. W. Mertens, Perceived size and distance of familiar objects, *Percept. mot. Skills*, 25, 1967, 213–225; W. C. Gogel, The effect of set on perceived egocentric distance, *Acta Psychol.*, 28, 1968, 283–292.

Method

Experiment I

Stimuli and apparatus. Two sizes of a playing card (the 10 of clubs) were viewed from 10 ft. away. One card was of normal size, the other was twice as large (*i.e.* simulated a card of normal size at 5 ft.). These stimuli were produced by transilluminating two positive transparencies and were observed monocularly (with the right eye) in an otherwise dark visual field. The luminance of the cards for *O* was .16 ftL., and the bottom of each of the cards (normal, or 'small,' and large) was at *O*'s eye level. There was a horizontal distance of 10 in. between the cards for both successive and simultaneous presentations.

Perceived size was measured by *O*'s adjustment of the lateral distance between two small posts located at about waist level. A meter stick attached to the apparatus permitted *E* to read the lateral separation between the posts from a position invisible to *O*. The observation position was totally dark and the size-indication apparatus was not visible to *O*. Great care was taken that no light except that from the stimuli was visible to *O* at any time.

Instead of the dark field with the cards, a visual alley with many cues (a calibration alley) could be presented. The calibration alley, about 24 ft. long × 3 ft. wide, had a floor formed by white cloth on which numbered rectangles were mounted at distances of 2, 5, 9, 14, and 20 ft. from *O*. The *O* viewed this full-cue alley binocularly from an observation position containing a head and chin rest at the front of the alley. The alley floor was 1 ft. below *O*'s eye level. Neither the calibration alley nor the numbered rectangles were visible during presentation of the cards. The purpose of the calibration alley was to determine that verbal reports of perceived distance were proportional to physical distances under full-cue conditions of observation. Since physical and perceived distance tend to be proportional in full-cue situations, such a result would validate the method of using verbal reports to measure the perceived distances of the playing cards under the conditions in which familiar size was the only distance cue available.[2]

Procedure. Three groups of 20 *O*s were used in the experiment. One group was first presented with the large card and then with the small card. There were approximately 25 sec. between presentations. With a second group of 20 *O*s, the procedure was the same except that the order of presenting the two sizes of cards was reversed. The third group of 20 *O*s was presented with both cards simultaneously.

Prior to the presentation of any of the cards, all *O*s were in total darkness for 10 min., in order to provide isolation from prior visual stimulation. The *O*s indicated the apparent distance of the cards by verbal reports in feet or inches or a combination of both. The perceived size (width) of the cards was indicated by means of the small posts at the observation position, *O* grasping one post in each hand and moving the right post laterally until the distance between the two posts seemed (kinesthetically) to be the same as the width of the card. This adjustment was made twice for each presentation of the card, and always *after* making the report of perceived distance.

After all the judgments of distance and size were completed for a particular

[2] Gogel, Hartman, and Harker, *loc. cit.;* Gogel and Mertens, *op. cit.*, 218.

O, verbal estimates of apparent distance were made with the full-cue alley. For this purpose, the lights in the full-cue alley were turned on and *O* moved to the appropriate observation position and indicated in feet or inches, or in a combination of both, the apparent distance from himself of each of the numbered rectangles. These judgments were requested in a random, prearranged order.

The *Os* were members of an introductory psychology class. All *Os* had a visual acuity of at least 20/20 in both eyes (as measured with an orthoscope), and none were acquainted with the purpose of the experiment.

Experiment II

Experiment II was the same as Experiment I except that the brightness of the playing cards was .70 ftL. and that the calibration alley was not used. A different group of 60 *Os* enrolled in the introductory psychology course was used. The acuity requirements were the same as for Experiment I.

Results

Since the results from Experiment I and II were very similar, the data from the presentations of cards in both experiments were combined and are summarized in Table I. It is also of importance that the average reports of distance from the calibration situation in Experiment I for the 2, 5, 9, 14, and 20 ft. distances were 1.6, 3.8, 6.9, 11.3, and 16.5 ft. respectively. It is thus clear that reported distance tended to be proportional to but less than the physical distance in the calibration alley, *i.e.* that the reported distances in Table I can be used to calculate perceived ratios even though they underestimate the distances by a multiplicative constant.

Discussion

Table I shows that the average reported sizes (widths) were greater, but not twice as great, for the large than for the small card. If familiar size determined perceived size, the large and small cards should have been reported to have the same width. If retinal size determined perceived size, the width of the large card should have

TABLE I
Average Reported Distance and Average Reported Size: Experiments I and II

	First presentation	Second presentation	Simultaneous presentation
Reported distance in ft.			
Small card	5.3	8.9	6.2
Large card	4.0	2.8	3.7
Reported size (width) in in.			
Small card	1.6	1.6	1.6
Large card	2.3	2.3	2.6

been reported to be twice as great as that of the small card. Thus it seems that *Os'* judgments of width were influenced both by the familiar and the retinal size of the card. From this result it is not clear how the judgments of width were related to judgments of distance. This problem will be examined in a later study.[3]

Table I also shows that the group of *Os* first presented with the small card reported it to be at a distance that was greater than the distance obtained from the group of *Os* first presented with the large card. The difference of 1.3 in. between these first prseentations was significant at the .05 level ($t_{78} = 2.45$). It follows that absolute retinal size was a significant determinant of perceived distance. Nevertheless, if the cue of absolute retinal size were a completely effective cue in determining perceived distance, the reported distance of the small card would have been twice that of the large card. Since the ratio was 1.3:1 rather than 2:1, it is clear that absolute size, although significant, was not a very effective cue to perceived distance.

If the reported distances in Table I had been physically correct, the ratios of the distances of the small card to the distances of the large card would have been 2:1 throughout the table. The ratios obtained from the data on distance in Table I were 1.3:1, 3.2:1, and 1.7:1 for the first, second, and simultaneous presentations respectively. If these ratios were constant, one could conclude that the same processes occurred in the judgments of distance for all three kinds of presentations. From the data, however, it is clear that additional or different factors were operative in the three cases. These results indicate that the different conditions of presentation effectively modified the perception of distance from familiar size and that they provide the basis for differentiating experimentally between the cues of absolute and relative size.

The relative cue was assumed to occur as a result of comparisons between different retinal sizes and to be most heavily weighted in the simultaneous presentations of the cards. The data support this assumption. The ratio of the reported distance of the small card as compared with the simultaneously presented large card (1.7:1) was sufficiently close to 2:1 that it can be concluded that for the simultaneous presentations, the ratio of perceived distances is similar to the ratio of retinal sizes. The difference between the obtained 1.7 and the hypothetical 2.0 probably can be attributed to the action of the equidistance tendency, which can reduce perceived depth between

[3] W. C. Gogel, The sensing of retinal size, *Vision Res.*, in press.

simultaneously presented objects.[4] It can be concluded tentatively that relative size as a cue to distance tends to result in ratios of perceived distance that are equal to the ratios of the retinal sizes of the familiar object.

It was also assumed that with the first presentations most weighted with the absolute cue, with the simultaneous presentations most weighted with the relative cue, and with the second presentations determined by both cue systems, the reported distances obtained from the second presentations could be explained in terms of the results from the first and simultaneous presentations. That explanation would be as follows: On the first presentation, *O* perceives the playing card to be at some particular distance as a result of the absolute cue. On the second presentation, *O*, shown a playing card one-half or twice the retinal size of the playing card presented first, uses the relative cue between the cards and perceives this second card at twice or one-half the distance (respectively) of the first card. This process can be expressed mathematically, where D' is perceived distance and θ is retinal size with the subscripts 1 and 2 referring to the first and second presentations respectively, as

$$D'_2 = D'_1\left(\frac{\theta_1}{\theta_2}\right). \qquad [1]$$

If Equation 1 is applied to the judged distances of Table I, it predicts that the results from the second presentations should be 8.0 for the small card and 2.6 for the large card. These values are close to the obtained values of 8.9 and 2.8. It should be noted that although Equation 1 specifies a way in which absolute and relative size cues can determine second (or following) presentations, the details of this equation are not essential to the distinction between absolute and relative size cues. For example, instead of concluding that a ratio of retinal sizes determines a ratio of perceived distances, one might consider the hypothesis that a ratio of retinal sizes determines a constant difference in perceived distance. The above equation is presented, however, because it more nearly fits the data obtained in the present study.

Table I shows that the difference between the judged distances from the second presentations was much larger than the difference between the judged distances from the first presentations. This

[4] W. C. Gogel, Equidistance tendency and its consequences, *Psychol. Bull.*, 64, 1965, 153–163.

change was significant at the .01 level ($t_{78} = 3.83$). It is as though a contrast effect occurred between the first and second presentations. A small difference in perceived distances between the first presentations resulted in a large difference in perceived distances between the second presentations. The proposed explanation in terms of absolute and relative size predicts this result. From Equation 1, whenever the ratio of the perceived distances from the first presentations is less than the ratio of the retinal sizes, the ratio of perceived distances from the second presentations must exceed the ratio of retinal sizes. It follows that under these conditions, the differences between second presentations are greater than the differences between first presentations. The demonstrated tendency for a change between first and second presentations to reflect a difference between absolute and relative cues can also be applied to distance cues such as accommodation, convergence, brightness, etc. The simplest application would be to use two different groups of *O*s and two different values of the distance cue being considered. Each of the two values of the distance cue would be presented successively to each of the groups with the order of presentation of the two values reversed for the two groups. If the perceived distance were more different between the groups on the second than on the first presentations, this would indicate that absolute and relative cues for that cue system are not identical. The importance of the distinction between absolute and relative cues is that psychophysical relations underlying these two kinds of cue systems are not necessarily identical, as the present study indicated. It is difficult to see how psychophysical relations having general applications can be obtained if these distinctions are ignored.

Summary

A differentiation between absolute and relative size as cues to distance was applied as an explanation to the difference in perceived egocentric distance which can occur between first and successive presentations of different retinal sizes of a familiar object. It was assumed that the absolute cue would be heavily weighted in the first presentation (a presentation following some time in the dark) and that the relative cue would be heavily weighted in the simultaneous presentation of the two retinal sizes. Both of these cue systems were assumed to operate in combination in determining the perceived distance associated with a second presentation of the familiar object. It was found that although in the first presentation,

on the average, the perceived distance of the familiar object increased with a decrease in retinal size, this increase was considerably less than that shown between second presentations. This change was predicted by considering the data resulting from both the first and simultaneous perceptions together. It was suggested that the method of using different groups and different orders of presentations would also be useful in differentiating between absolute and relative cues in other cue systems than that of familiar size.

ADAPTATION TO CHANGE IN MULTIPLE PROBABILITY TASKS

By David A. Summers, University of Kansas

Following Brunswik's proposition that man functions as "an intuitive statistician," several investigators have shown that *Ss* can effectively utilize multiple, probabilistic cues in order to make inferences about an interval-scaled criterion variable.[1] Most multiple probability-learning (MPL) studies, however, have involved *stationary* task characteristics; *i.e.* the stimulus-cue weights have remained more or less constant over trials. In contrast, *S*'s capacity to adapt to a change, or *shift*, in MPL task characteristics has received only limited attention. In this regard, it has been found that when the cue weights are reversed after *Ss* reach asymptote, adaptation to the new weights is slower than adaptation to the initial (preshift) weights.[2]

Although it has been demonstrated that *Ss* can indeed adapt to a change in MPL tasks, it is important to note that there are several different ways in which a change can occur. As pointed out elsewhere, MPL tasks can require *S* to discover either (1) the relevant (non-zero weighted) cue(s), (2) the regression function (*i.e.* rule) relating the cue(s) to the criterion variable, or (3) both the relevant cue(s) and rule.[3] Given that the relevant cue(s) and the task rule constitute distinct aspects of MPL problems, at least three types of task shift are possible. Specifically, the relevant cue(s) can change, while the regression function relating the cue(s) to the criterion remains invariant. Such a task change can be termed a *cue shift*. Conversely, the regression function can change, while the relevant cue(s) remains the same. Such a change

* Received for publication October 16, 1968. The research reported here was undertaken in the Group Effectiveness Research Laboratory, University of Illinois, and was supported by the Advanced Research Projects Agency, ARPA Order No. 454, Contract NR-177-472, NONR 1834(36).

[1] E. Brunswik, *Perception and the Representative Design of Psychological Experiments*, 1956. See also C. R. Peterson and L. R. Beach, Man as an intuitive statistician, *Psychol. Bull.*, 68, 1967, 29–45.

[2] C. R. Peterson, K. R. Hammond, and D. A. Summers, Multiple probability learning with shifting weights of cues, this Journal, 4, 1965, 660–663.

[3] D. A. Summers, Rule versus cue learning in multiple probability tasks, *Proc. 75th Annu. APA Conv.*, 1967, 43–44.

can be termed a *rule shift*. Finally, both the relevant cue(s) and the task rule can change; hence, a *complete shift*. While these distinctions concerning types of task shift have conceptual support, it remains to be shown whether they are empirically justified. The present study, therefore, concerns the effects of cue, rule, and complete shifts upon performance in a three-cue, probability-learning task.

Method

Design and subjects. In order to assess the effects of a rule shift, two types of linear task rules were utilized. Specifically, half of the *Ss* initially learned (in the preshift trials) to make inferences on the basis of a positive regression rule, while the remaining *Ss* initially learned a negative regression rule. Systematically varied in the experiment, therefore, were (*a*) type of rule involved in the preshift trials (positive vs. negative) and (*b*) type of task shift (No Shift vs. Cue Shift vs. Rule Shift vs. Complete Shift). The experiment thus took the form of a 2×4 factorial design with repeated measures over blocks of trials. A total of 120 University of Illinois undergraduates served as *Ss*.

Preshift task. Both the positive and negative tasks consisted of three cues, one of which accounted for all the systematic criterion variance. Values for each cue in the 60 preshift trials were obtained by selecting 60 digits (1–10) from a random-number table; cue intercorrelations were thus near zero. The criterion values for the positive task were generated by combining these cue values according to the following rule: $Y = X_1 + 0(X_2) + 0(X_3)$. The task was made probabilistic by randomly performing one of the following operations with regard to the Y values: +2, +1, +1, 0, 0, 0, −1, −1, −2. These operations yielded a task in which the final criterion values (Y') correlated .94 with X_1, and approximately zero with X_2 and X_3. The criterion values for the negative task were generated by combining the cue values as follows: $Y = 11 - X_1 + 0(X_2) + 0(X_3)$. Thus, when the relevant cue (X_1) had values of 1, 2, 3, etc., the criterion values (Y) were 10, 9, 8, etc., respectively. The same random-error operations performed in the positive task were also performed here. As a result, the final criterion values (Y') correlated −.94 with X_1 and approximately zero with X_2 and X_3.

Postshift task. The stimulus-cue values for the 60 postshift trials were likewise obtained from a random-number table; again, cue intercorrelations were near zero. For *Ss* in the No Shift (control) condition, the criterion values for the postshift trials were generated in exactly the same manner as generated in the preshift trials. In the Cue Shift condition, the postshift criterion was generated according to the same rule (positive or negative) utilized in the preshift trials, but by a different cue. In the Rule Shift condition, the rule by which the criterion was generated was reversed. Specifically, if the preshift criterion was a positive function of X_1, the postshift criterion was generated as a negative function of the same cue. Similarly, if the preshift criterion was a negative function of X_1, the postshift criterion was a positive function of that cue. In the Complete Shift condition, not only was the postshift regression rule reversed, but also a different cue served to generate the criterion.

These shift conditions can be placed in better perspective if they are related to parallel efforts to study task shifts in the traditional concept-learning framework.[4] Specifically, the rule shift is analogous to a *reversal* shift in concept-learning studies; *i.e.* in both cases the relevant stimulus dimension remains the same, while the rule that should be applied to that stimulus dimension is reversed. Similarly, the cue shift is analogous to a *nonreversal* shift in concept-learning studies; *i.e.* in both, only the relevant stimulus dimension is shifted. In all postshift conditions, the same random-error operations performed in the preshift trials were performed with respect to the postshift criterion. As a result, the limits of accuracy in the pre- and postshift trials were nearly identical.

Apparatus. Stimuli were presented on a 3 × 4 ft. plywood board on which were displayed three vertical 10-point scales. These scales represented the three cues. Directly below these vertical scales was a 10-point horizontal scale representing the criterion variable.

Instructions and procedure. *S*s were instructed (*a*) that the criterion was related to only one of the three cues present, (*b*) that the criterion was either a positive linear or negative linear function of the relevant cue, and (*c*) that a task change *might* occur. In addition, *S*s were informed that random variance prevented perfect accuracy. Each of the 120 trials consisted of the following steps: (1) *E* specified a value for each of the stimulus cues by circling the appropriate column numbers; (2) *S*s received 10 sec. in which to predict the criterion value; and (3) *E* indicated the correct criterion value by circling a number on the horizontal criterion scale.

Results

Performance was assessed by correlating *S*'s judgments with the criterion values (Y') over each block of 15 trials. The resulting accuracy correlations (r_a) were then transformed into Fisher's Z coefficients for subsequent analysis. Only *S*s whose performance on the final block of preshift trials exceeded chance ($r_a \geq .51$) were included in the study.[5] As a result, the mean accuracy correlation (over all *S*s) for the final block of preshift trials was substantial ($\bar{r}_a = .78$). Moreover, in order to assure that the treatment groups did not differ significantly immediately prior to the shift, accuracy correlations derived from the final preshift block were subjected to a 2 (Positive vs. Negative Rule) × 4 (No Shift vs. Rule Shift vs. Cue Shift vs. Complete Shift) analysis of variance. This analysis revealed no significant effects. In this respect, it should be noted that mean accuracy correlations for the four treatment conditions were nearly identical (±.04) on the last block of preshift trials.

[4] For example, see H. H. Kendler and T. S. Kendler, Vertical and horizontal processes in problem solving, *Psychol. Rev.*, 69, 1962, 1–16.

[5] Seven *S*s failed to learn the preshift solution, and were dropped from the experiment. Additional *S*s were obtained to complete the task in their place.

Accuracy correlations derived from each block of postshift trials were treated according to a 2 (Positive vs. Negative Rule) × 4 (No Shift vs. Rule Shift vs. Cue Shift vs. Complete Shift) analysis of variance with repeated measures over four blocks of trials. This analysis yielded three statistically reliable effects. Specifically, accuracy varied according to blocks, F (df = 3,336) = 120.63, $p <$ .001; and according to shift condition, F (df = 3,112) = 16.54, $p <$ .001. In addition, a significant Blocks × Shift interaction was found, F (df = 9,336) = 13.48, $p <$ .001. These findings are shown graphically in Fig. 1. As can be seen, the blocks main effect can be attributed to a significant improvement over trials; *i.e.* over all conditions, accuracy was substantially greater for the last block of trials ($\bar{r}_a$ = .83) than the first ($\bar{r}_a$ = .52). With regard to the main effect due to the shift conditions, individual comparisons revealed that performance in the No Shift condition ($\bar{r}_a$ = .88) was significantly ($p <$ .05) better than that observed in the Cue ($\bar{r}_a$ = .80) and Rule Shift ($\bar{r}_a$ = .81) conditions, while performance in the latter two conditions was significantly ($p <$ .01) better than in the Complete Shift condition ($\bar{r}_a$ = .59).[6] The Blocks × Shift interaction can be accounted for by differences in performance on Block I of the postshift trials. Specifically, *S*s in the No Shift condition were significantly ($p <$ .01) more accurate than were *S*s in either the Rule, Cue, or Complete Shift conditions. Furthermore, while performance in the Rule and Complete Shift conditions did not differ, *S*s in both conditions were significantly ($p <$ .05) *less* accurate on Block I than were *S*s in the Cue Shift condition.

Discussion

The findings reported here indicate that adaptation to change in a multiple-cue, probabilistic task is substantially affected by *which* aspects of the task undergo change. In the present study, both (a) the initial decrement in performance and (b) the rate of improvement following the shift were affected by the type of shift involved. With regard to (a), the initial decrement in performance was least when S was required only to utilize a different cue in order to achieve inductive accuracy. As Fig. 1 shows, the decrement in performance was substantially greater when S was required to utilize a different induction rule (*i.e.*

[6] All significance levels for the individual comparisons were established by the Newman-Keuls procedure; see B. J. Winer, *Statistical Principles in Experimental Design,* 1962.

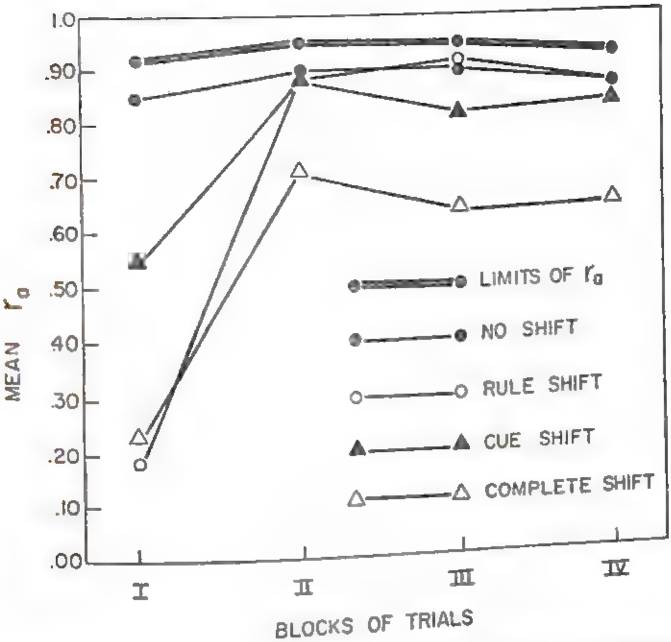


FIG. 1. MEAN ACCURACY CORRELATIONS (r_a) BY SHIFT CONDITION OVER POSTSHIFT TRIALS

a Rule or Complete Shift). In certain respects, the initial difference in performance between the Cue and Rule Shift conditions is not surprising. Inasmuch as the Rule Shift involved a direct reversal of the preshift rule, we would expect *Ss'* initial postshift responses to be maximally different from the criterion; *i.e.* the correct postshift responses were inversely correlated with the responses associated with the preshift solution. In contrast, we would *not* expect the initial postshift responses in the Cue Shift condition to be maximally different from the criterion. Inasmuch as the stimulus-cue values had near-zero intercorrelations, the correct postshift responses were simply uncorrelated with the responses associated with the preshift solution. In short, the initial difference in performance between the Cue and Rule Shift conditions can be attributed to the particular type of rule shift studied here, *i.e.* a direct rule reversal.

With regard to (*b*), it is important to note that *Ss* in both the Rule and Cue Shift conditions approached the limits of accuracy quite rapidly, *i.e.* by the end of the second block of postshift trials. On the other hand, *Ss* in the Complete Shift condition remained substantially below the limits of accuracy even

at the conclusion of the 60 postshift trials. One possible explanation of these findings can be derived from an hypothesis-testing framework. In this regard, let us assume that when the solution change was encountered, the *Ss* initially hypothesized that only a single aspect of the task had changed. If such were the case, the *Ss* in the Cue and Rule Shift conditions were faced with a narrow range of alternatives; *i.e.* only three cues were present, and they had been instructed that only two task rules (positive linear or negative linear) were possible. As a result, *Ss* in the Cue and Rule Shift groups could achieve the correct solution by testing only three hypotheses: (1) that the preshift rule had been reversed; (2) that cue X_2 was now relevant; and (3) that cue X_3 was now relevant. On the basis of the number of alternatives available to *Ss* in the Rule and Cue Shift conditions, one might therefore have expected these *Ss* to adapt to the change easily, and within approximately the same number of trials. And, as can be seen in Fig. 1, such was indeed the case. The situation in the Complete Shift condition was somewhat different. Clearly, the *Ss* in this condition could not have achieved the correct solution by testing only the three hypotheses outlined above. Rather, these *Ss* were required to test two *additional* hypotheses: (1) that cue X_2 was now relevant *and* the preshift rule had been reversed; and (2) that cue X_3 was now relevant *and* the preshift rule had been reversed. Inasmuch as the Complete Shift *Ss* were confronted with a larger number of alternatives, it is not surprising that these *Ss* were slower in their adaptation to the task change.

The interpretations advanced here can be regarded as only tentative, however. In particular, support for the hypothesis-testing interpretation of performance in the present situation would require additional evidence from shifting MPL tasks involving (*a*) a different number of stimulus cues and (*b*) different (*e.g.* nonlinear) induction rules.[7]

[7] For example, see K. R. Hammond and D. A. Summers, Cognitive dependance on linear and nonlinear cues, *Psychol. Rev.*, 72, 1965, 215–224.

EFFECTS OF THREE SITUATIONAL VARIABLES AND TYPE OF REINFORCEMENT ON DELAY PREFERENCE

By Thomas A. Glass, Hawaii State Hospital

The tendency to forgo immediate reinforcement in favor of delayed reinforcement has received considerable attention in the literature. Preference for delayed reinforcement has been related to various personality characteristics such as achievement motivation and intelligence, as well as to group-membership correlates such as age and socioeconomic status. By contrast, the influence of situational variables on relative preference for immediate and delayed outcomes has not been examined so thoroughly. Most of the studies in this area have been concerned solely with relative preference for immediate and delayed *rewards*. Two such studies by Lipsitt and Castaneda and by Mischel and Metzner have demonstrated that children's preferences for delayed rewards were negatively related to the length of the delay interval.[1] Mahrer has found that both a child's trust in the examiner and his expectancy that a delayed reward would be forthcoming were positively related to preference for delayed rewards;[2] and Metzner has found, by varying the nature of the delay period, that children were more likely to choose a delayed reward if they could work for it than if they could only wait for it.[3] Of the relatively few recent studies of preference for immediate vs. delayed punishments, D'Amato and Gumenik have found that human Ss preferred an immediate shock to a delayed shock when the delayed

*Received for publication October 10, 1968. The article is based on a dissertation submitted to the Graduate College, University of Nebraska, in partial fulfillment of the requirements for the Ph.D. in psychology. The author would like to express his appreciation to David Levine, his dissertation advisor, and to John C. Loehlin for his guidance in conceiving and designing the study.

[1] L. R. Lipsitt and A. Castaneda, Effects of delayed rewards on choice behavior and response speeds in children, *J. comp. physiol. Psychol.*, 51, 1958, 65–67; W. Mischel and R. Metzner, Preference for delayed rewards as a function of age, intelligence, and length of delay-interval, *J. abnorm. soc. Psychol.*, 64, 1962, 425–431.

[2] A. R. Mahrer, The role of expectancy in delayed reinforcement, *J. exp. Psychol.*, 52, 1956, 101–106.

[3] R. Metzner, Work and delay of gratification, unpublished doctoral dissertation, Harvard University, 1962.

shock followed a randomly varied time interval, and Cook and Barnes substantiated these findings by demonstrating that human *Ss* tended to choose a short rather than a long delay of inevitable shock; the longer the delay, the less tendency there was to choose it.[4] Finally, two recent studies are unique in that their experimental designs provide for a direct comparison of preference for both delayed and immediate rewards and delayed and immediate 'punishments.' Mischel and Grusec found that as the probability of receiving a delayed outcome increased, children were more likely to choose delayed rewards and immediate 'punishments.'[5] In a related study Grusec found that as the size of delayed outcomes increased, children made more delayed-reward and more immediate-'punishment' choices.[6]

The present study sought to investigate the effects of three situational variables on the relative preference of adults for delayed outcomes in both reward and 'punishment' situations. Specifically, the three situational variables selected were *certainty* that delayed rewards and 'punishments' would actually be received, *pleasantness* of the waiting period prior to receiving delayed rewards and 'punishments,' and *vividness* with which delayed rewards and 'punishments' were presented. Levels of certainty about receiving delayed outcomes were varied to better understand the influence of this variable on delay preference in an adult population; selection of this variable followed from the earlier studies with children which demonstrated that two important variables in affecting delay preference are the level of expectancy about receiving a delayed outcome[7] and the probability of occurrence of a delayed outcome.[8] Pleasantness of the delay period was manipulated to clarify its role in influencing delay preference in adults; this followed from Metzner's finding that the nature of the waiting period prior to delayed reinforcement significantly affected delay preference in children.[9] Vividness was also varied, for although little attention has been given either

[4] M. R. D'Amato and W. E. Gumenik, Effects of immediate vs. randomly delayed shock on instrumental response and cognitive processes, *J. abnorm. soc. Psychol.*, 60, 1960, 64–67; O. J. Cook and L. W. Barnes, Choice of delay of inevitable shock, *J. abnorm. soc. Psychol.* 68, 1964, 669–672.

[5] W. Mischel and J. Grusec, Waiting for rewards and punishments: Effects of time and probability on choice, *J. Pers. soc. Psychol.*, 5, 1967, 24–31.

[6] J. Grusec, Waiting for rewards and punishments: Effects of reinforcement value on choice, *J. Pers. soc. Psychol.*, 9, 1968, 85–89.

[7] Mahrer, *loc. cit.*

[8] Mischel and Grusec, *loc. cit.*

[9] Metzner, *loc. cit.*

to the manner in which reinforcements are presented or to the possible effects that varying presentation conditions may have on delay preference, the vividness of reinforcement presentation may be potentially useful in clarifying the requisites of delay preference. In the commercial promotion of consumer goods, for example, much value is attributed to vividness of presentation in packaging or display as a means of enhancing preference and influencing consumers' decisions.

Method

Apparatus. Ss participated individually in a small laboratory room and were seated in front of a vertical panel situated on a table. The panel, about 2 ft. high above the table top and approximately 3 ft. wide, had a small light at the top and a scoreboard in the center. The scoreboard was covered by a hinged cover which was kept closed except when *S* was permitted to see his score. On the table at the base of the panel were four telegraph keys, each marked with a symbol representing one of the four suits of a standard deck of playing cards. When any of the keys was pressed, the click of a closing relay could be distinctly heard. The light at the top of the panel could be flashed by *E*. A buzzer behind the panel sounded in conjunction with the flashing light. On a table next to the panel and in full view of *S* were a tape recorder and several 'electronic' devices covered with dials, pilot lights, and switches. During the experiment *S* was seated in front of the panel and *E* was seated behind the panel out of *S*'s sight.

Procedure. Ninety-six undergraduates (men) served as *Ss*. Tape-recorded instructions informed *S* that the experiment dealt with "subconscious intuition" and that he was to guess the correct suit of cards being turned one by one by *E* out of sight behind the panel. This format provided *S* with a task about which he presumably had few preconceived notions as to his level of skill. *S* was told to indicate his responses by pressing one of the four keys in front of him and that his responses would be "electronically" recorded with points being awarded for each correct response. These points were to accumulate automatically on the scoreboard during each series of trials. The scoreboard was, however, to remain covered except at various specified intervals between trial series when *S* would be permitted to see his score. Unknown to *S*, scores appearing on the scoreboard were inserted by *E* from behind the panel during each trial series. The scores presented were based on a predetermined reinforcement schedule and were the same for all *Ss* in a given experimental group. In fact, the accuracy or inaccuracy of *S*'s responses in the card-guessing task was unrelated to his score. At the completion of each trial series, *S* was told to check his score on the scoreboard and record it on a summary sheet.

Following a short practice trial to familiarize himself with the apparatus and procedure, each *S* played two experimental card-guessing games. Each game consisted of four sequences of card-guessing followed by reinforcement, i.e. by *S*'s receiving a score. In one of these games *S* received his score upon completing each series of trials (immediate reinforcement), while in the other game *S* had to wait 2 min. after each series of trials to receive his score

(delayed reinforcement). In comparing their scores with a fictitious group mean, one-third of the Ss experienced favorable outcomes, one-third unfavorable outcomes, and one-third neutral outcomes. The delayed-reinforcement game was further varied in such a way that half the Ss played under conditions of certainty about the outcome (reinforcement followed each trial series) and half under conditions of uncertainty (reinforcement followed randomly only 50% of the trial series). Furthermore, for half the Ss reinforcement was presented in a vivid manner (score presentation was accompanied by a flashing panel light, sounding buzzer, and poker-chip payoff), while half received reinforcement in a nonvivid manner (only score was shown). The delay period for half the Ss was pleasant (reading aloud an interesting magazine article), and for half unpleasant (reading aloud columns of numbers). Experimental conditions for the immediate-reinforcement game were the same for all Ss in that reinforcement was always uncertain and nonvivid. Ss were randomly assigned to each of the experimental treatments, and order of game presentation was counterbalanced within each group.

At the completion of the experiment S was asked to complete a questionnaire designed to indicate which of the two games he preferred. At this time, E also informally interviewed S to determine S's naïveté. Subjects were considered naïve if, in E's judgment, they believed the scores they had received resulted from their own performance in the guessing task. Of 136 Ss in the initial population pool, only three had to be dropped because they seriously questioned the assertion that their scores really represented their card-guessing skill.

Results and Discussion

Game-preference scores, reflecting Ss' relative preferences for the immediate- or delayed-reinforcement game, constituted the principal dependent variable. Game-preference scores ranged from 0 through 9, 0 representing the greatest degree of preference for immediate reinforcement (low delay preference) and 9 representing the greatest degree of preference for delayed reinforcement (high delay preference). Table I represents the game-preference scores for all 96 Ss. Each entry is the mean for the four Ss in one of the 24 experimental subgroups (three outcome conditions—favorable, unfavorable, neutral—each with a certainty and uncertainty subgroup; each of these subgroups with a vivid and nonvivid subgroup; each of these vivid and nonvivid subgroups with a pleasant and unpleasant delay period) in the delayed-reinforcement game. Each of these Ss had, of course, also played the immediate-reinforcement game under one of its three outcome conditions—favorable, unfavorable, neutral.

Table II shows the summary of analysis of variance of these mean game-preference scores. Bartlett's test of homogeneity of variance was applied to the variances associated with each of

TABLE I

MEAN GAME-PREFERENCE SCORES

	Favorable outcome		Unfavorable outcome		Neutral outcome	
	Pleasant	Unpleasant	Pleasant	Unpleasant	Pleasant	Unpleasant
Certain						
Vivid	4.25	3.75	4.50	1.50	5.75	.50
Nonvivid	8.75	6.00	7.25	2.25	4.50	5.25
Uncertain						
Vivid	4.75	3.00	1.75	3.25	6.00	.75
Nonvivid	1.25	2.50	2.00	2.50	.75	2.25

Note: Game preference scores ranged from 0 through 9 with means above 4.50 representing preference for the delayed-reinforcement game.

the three reinforcement groups, and the variances proved to be satisfactorily homogeneous ($\chi^2 = .51$). The data in Table II indicate that certainty did account for a significant amount of variance ($F = 9.10$). The mean game-preference score for the 48 *Ss* who experienced the certainty conditions was 4.52, while the mean game-preference score for the 48 *Ss* in the uncertainty conditions was 2.56. Furthermore, the pleasantness variable also reached significance ($F = 5.34$) with a mean game-preference score of 4.29 for the 48 *Ss* in the pleasantness conditions as compared to 2.79 for the 48 *Ss* in the unpleasantness conditions. The vividness variable did not reach significance, and its effects on delay preference are less clear. Vividness may exert an in-

TABLE II

ANALYSIS OF VARIANCE OF GAME-PREFERENCE SCORES

Source	*SS*	*df*	*MS*	*F*
Reinforcement group (R)	26.40	2	13.20	1.30
Certainty (C)	92.04	1	92.04	9.10**
Vividness (V)	5.04	1	5.04	.49
Pleasantness (P)	54.00	1	54.00	5.34*
R × C	8.77	2	4.39	.43
R × V	3.27	2	1.64	.16
R × P	5.06	2	2.53	.25
C × V	80.67	1	80.67	7.97**
C × P	30.37	1	30.37	3.00
V × P	18.37	1	18.37	1.81
R × C × V	11.40	2	5.70	.56
R × C × P	23.69	2	11.84	1.17
R × V × P	67.69	2	33.84	3.34*
C × V × P	8.17	1	8.17	.80
R × C × V × P	6.39	2	3.20	.31
Within (replication)	728.50	72	10.12	
Total		95		

* $p < .05$.
** $p < .01$.

fluence only to the extent that it interacts with pleasantness or certainty. It is of interest that the type of reinforcement (*i.e.* favorable, unfavorable, neutral) did not account for a significant amount of variance.

It appears, then, that certainty of delayed outcomes and pleasantness of delay period significantly increase preference for both delayed rewards and 'punishments.' Findings for the favorable-reinforcement group are consistent with those that have been reported in that increased certainty of a delayed reward increased preference for such outcome. However, the finding that unfavorable-reinforcement *S*s also indicated increased preference for delayed 'punishment' when the delayed outcome became more certain seems at variance with the Mischel and Grusec finding that children were more likely to prefer immediate 'punishment' as the probability of delayed reinforcement increased.[10] Since the unfavorable outcomes in the present study were not very severe, it is possible that the adult *S*s were more desirous of being certain that they would receive feedback about their performance, even when they anticipated a negative outcome. Direct comparison of the two studies is, however, confounded by substantial differences in the experimental designs. In addition to differences in the age of *S*s, the nature of the negative reinforcements differed, as did length of delay. For example, whereas *S*s in the present study experienced 2-min. delays, the reinforcement delays for Mischel and Grusek's *S*s varied from one day to one month.[11] By systematically varying delayed-outcome probabilities and delay-period activities and observing their effects on delay preference in various age groups, subsequent research should serve to clarify the present findings.

Summary

To investigate the influence of three situational variables and type of reinforcement on delay preference, a card-guessing game was administered to 96 university students (men) who were then given favorable, unfavorable, or neutral scores. In one game *S*s saw their scores at once (immediate reinforcement); in another game *S*s waited 2 min. before receiving their scores (delayed reinforcement). The games were played under various conditions

[10] Mischel and Grusec, *loc. cit.*
[11] Mischel and Grusec, *loc. cit.*

of (1) certainty about receiving a score, (2) vividness of score presentation, and (3) pleasantness of the waiting period preceding score presentation. Relative preference for one of the two games constituted the dependent variable. Whereas certainty and pleasantness increased preference for both delayed rewards and 'punishments,' vividness effects were more difficult to interpret.

THE EFFECT OF TYPE SIZE AND CASE ALTERNATION ON WORD IDENTIFICATION

By Frank Smith, Deborah Lott, and Bruce Cronnell,
Southwest Regional Laboratory, Inglewood, California

The distinctive-feature theory of letter recognition proposed by Gibson[1] has been developed to account not only for single-letter identification but also for the identification of words and other sequences of letters.[2] The present paper proposes that a reader does not normally attend to individual letters but rather to sets of features, various combinations of which may represent the same word. This view is essentially a 'whole word' approach to word recognition,[3] except that it does not demand familiarity with the total outline or configuration of a word. Instead, it proposes that the reader (1) extracts distinctive-feature information simultaneously from several parts of the configuration and (2) integrates this information for identification of the whole. The distinctive features mentioned are defined as elements or properties of letters or of groups of letters (including words), the discrimination of which reduces the set of alternatives that the total configuration, letter or word, might be. Any minimum combination of features sufficient to determine uniquely a particular letter or word is termed a criterial set of features. Sets of features are termed equivalent when more than one feature pattern represents a single letter or word. For example, A, a, and *a* comprise quite different sets of features that are equivalent be-

* Received for publication October 24, 1968. The study reported is part of research conducted pursuant to a contract between the Southwest Regional Laboratory and the United States Office of Education. Dr. Smith is now at the Ontario Institute for Studies in Education, 102 Bloor Street West, Toronto 5, Ontario, Canada.

[1] Eleanor J. Gibson, J. J. Gibson, A. D. Pick, and H. Osser, A developmental study of the discrimination of letter-like forms, *J. comp. Physiol*, 55, 1962, 897–906; Eleanor J. Gibson, Learning to read, *Science*, 148, 1965, 1066–1072.

[2] Frank Smith, The visual identification of words, unpublished doctoral dissertation, Harvard University, 1967; The use of featural dependencies across letters in the visual identification of words, *J. verb. Learn. verb. Behav.*, in press; Familiarity of configuration vs. discriminability of features in the visual identification of words, *Psychon. Sci.*, 14(6), 1969, 261–262.

[3] I. H. Anderson and W. F. Dearborn, *The Psychology of Teaching Reading*, 1952, 191–193.

cause they all represent the same letter, just as HAT, hat, and *hat* represent the same word.

The model proposed asserts that unfamiliarity with a total word form should not affect word identification adversely unless discrimination of features within the word is impaired. There is, however, evidence apparently contradictory to this view. Anderson and Dearborn report that material printed in alternating upper and lower case (see Condition 6 in Fig. 1) is more difficult to read than either all upper case (Condition 1) or all lower case (Condition 2).[4] These results could be due to readers' inability to treat upper- and lowercase letters as 'equivalent' when they occur alternately within a word; in which case, the model here proposed would not be upheld. But the Anderson and Dearborn results could also be attributed to the fact that the relative size of elements—which is a cue to the discrimination of lowercase letters—is interfered with when these letters are mixed with the taller capitals. The present study accordingly predicted that alternation of upper- and lowercase letters would not interfere with the identification of words when the size of the capitals was reduced, even though this still resulted in a configuration unfamiliar to the reader (Condition 3); but that variation in the size of alternate letters would interfere with word discrimination, even though all the letters were in lower case (Condition 5). It predicted that variation in the size of alternate letters would *not* affect word identification in all-capital text because relative size would not appear to be a cue for discrimination of uppercase letters (Condition 4). The predictions were tested in an earlier study by measuring the length of time taken by *Ss* to read a passage printed in the six

RELATIVE SIZE CUE	CASE	EXAMPLE
Irrelevant	Upper	1. THAT WE ARE AS YET QUITE IGNORANT OF EVEN THE SIMPLEST
Maintained	Lower	2. That we are as yet quite ignorant of even the simplest
	Mixed	3. That we ARE as yet QUITE IGNORANT OF even the SIMPLEST
Irrelevant	Upper	4. THaT wE ARe As YeT QUiTe IGNoRANT OF EvEN ThE SIMPLEST
Disrupted	Lower	5. That we are as yet quite ignorant of even the simplest
	Mixed	6. ThAt We ArE As YeT qUiTe IgNoRaNt Of EvEn ThE SiMpLeSt

FIG. 1. EXAMPLES OF SIX TYPOGRAPHIC CONDITIONS: SET-1 BOOKLETS
(Line widths approximately equal)

[4] Anderson and Dearborn, *loc. cit.*

typographic styles of Fig. 1.[5] All the predictions were supported. But although no special instructions were given with respect to comprehension, and although *Ss* read with normal intonation, the possibility remained that the results of that study were not wholly attributable to the ease of identification of individual words. In the present study, therefore, *Ss* were presented with a purely search-and-identification task requiring neither 'comprehension' of the text nor verbalization of response.

METHOD

Materials. The six typographic conditions were presented in booklets comprised of six 150-word passages of text. In half of the booklets (Set 1), line width was controlled (by varying letter width) so that equivalent groups of words occupied approximately the same size line (as indicated in Fig. 1). In the remainder of the booklets (Set 2), the width of type was not controlled to equalize comparable lines (see Fig. 2). Apart from the difference due to line and letter width, the two sets were identical. Each of the six passages in every booklet was printed in one of the six typographic styles, with all styles represented in all booklets. The order of passages was constant, but the order of conditions was determined according to a Latin-square design.[6] No condition occurred more than once in any one serial-order position, and no condition preceded or followed any other condition more than once.

Subjects and procedure. A total of 216 college students served as *Ss*, half receiving Set-1 and half receiving Set-2 booklets. Testing was conducted in group sessions, the 36 copies of each version of the test booklet being allocated at random.

For each of the six passages *Ss* were given a list of 20 words. Each word was between four and eight letters long, had occurred only once in the passage, and was not included in any other list. Passages and lists were printed on facing pages of test booklets: text on the left and word list on the right. Each

RELATIVE SIZE CUE	CASE	EXAMPLE
Irrelevant	Upper	1. THAT WE ARE AS YET QUITE IGNORANT OF EVEN THE SIMPLEST
Maintained	Lower	2. That we are as yet quite ignorant of even the simplest
	Mixed	3. ThAt we ArE as yet QuitE IGNOrANT OF eveN tHE siMpLest
Irrelevant	Upper	4. THAT WE ARE AS YET QUITE IGNORANT OF EVEN THE SIMPLEST
Disrupted	Lower	5. That we are as yet quite ignorant of even the simplest
	Mixed	6. ThAt We ArE As YeT qUiTe IgNoRaNt Of EvEn ThE SiMpLeSt

FIG. 2. EXAMPLES OF SIX TYPOGRAPHIC CONDITIONS: SET-2 BOOKLETS
(Width of letters unchanged)

[5] Smith, *op cit.*, *Psychon. Sci.*, 14(6), 261–262.
[6] E. F. Lindquist, *Design and Analysis of Experiments in Psychology and Education*, 1953, 258 ff.

line of text was numbered. *Ss* were asked to find the listed words in the passage and to write in a box beside each test word the number of the line of text in which it occurred. A short trial passage was given to assure understanding of the instructions. To discourage *Ss* from reading straight through the passage, words in the list were presented in random order, the same for all *Ss*. To minimize any tendency to attempt 'visual matches,' word lists were printed at random in lower, upper, and mixed case italic rather than in the Gothic used in the passages. The score for each condition was the number of words correctly identified by *Ss* during 2½ min.

Results

There was no significant difference between the two sets of booklets, and the data for both sets were combined (Table I). Mean scores for the six conditions supported the prediction that there would be no reduction in *Ss*' ability to identify words in which case was alternated if the size was held constant (Condition 3); but there was a significant reduction in ability to identify words in which both size and case were alternated (Condition 6). The prediction that interference with the relative size of elements of lowercase print (Condition 5) would reduce ability to identify words was also supported. However, interference with relative size in uppercase print (Condition 4) also resulted in a significant reduction in *Ss*' ability to identify words, contrary to the prediction that this condition would be no more difficult than normal uppercase print (Condition 1).

Analysis of variance for repeated measures[7] showed a significant difference at the .01 level between typographic conditions, with neither the set variable nor the interaction between sets and typographic conditions significant. Tukey's test for the significance of differences between all possible pairs of means showed no significant difference between the three conditions in which relative size of the letters was held constant and no significant difference between the three conditions in which relative size of the letters

TABLE I

Mean Number of Words Correctly Identified in Each Typographical Condition: Data from Both Sets of Booklets Combined

	Condition					
	1	2	3	4	5	6
Mean number of words correctly identified	10.98	11.13	11.06	10.03	10.19	10.05

[7] B. J. Winer, *Statistical Principles in Experimental Design*, 1962, 302–309.

was distorted; but each of the three constant-size conditions was significantly different at the .01 level from the three mixed-size conditions. These differences held whether the two sets of data were combined or considered individually.

Discussion

The results of this experiment support the hypothesis that disruption of 'total word form' does not interfere with the identification of words unless discriminability of elements is disrupted, *e.g.* by mixing the size of the individual letters. Nor is there any indication that particular groups of letters ('spelling patterns') are identified on the basis of familiar configurations. The only unpredicted finding was that size alternation of uppercase letters interfered with a search-and-identification task although it did not interfere with the earlier reading task when some comprehension and verbalization of response were involved. One possible explanation for this conflicting result lies in the difference in relative response complexity when *S* is searching for a word compared with when he is involved with a continuous reading task. During a reading task, there is only one correct response that *S* can make to any stimulus, *i.e.* he can say "transfer" whether the stimulus is TRANSFER, TrAnSfEr, or tRaNsFeR. In the search task, however, *S* looking for "transfer" in the all-capitals, mixed-size condition does not know whether his 'response' should be TrAnSfEr or tRaNsFeR; and it has frequently been demonstrated that response uncertainty is more difficult to cope with than stimulus uncertainty.[8] If the differences in the present study are primarily attributable to a size effect, as predicted by the hypothesis, the interference should be apparent in Conditions 4, 5, and 6. If the difference is related to the counter hypothesis, *i.e.* to an effect of unfamiliarity with the entire configuration, then the difference should also be apparent in Condition 3. Condition 3 was not any different from Conditions 1 and 2, so the relative difficulty of Condition 4, while not expected, is not considered damaging to the experimental hypotheses.

The most significant aspect of the results is perhaps not that *S*s experience about 10% more difficulty in certain conditions, but that they are able to handle the more unusual conditions at all. Experienced readers adapt quickly to all distortions of type,

[8] W. R. Garner, *Uncertainty and Structure as Psychological Concepts*, 1962, 35–38.

just as they can adapt to most forms of handwriting. Many of the various typographic and calligraphic arrangements of words and letters differ so radically that they cannot be viewed as 'transformations' but, we submit, must be considered as alternative forms. As the present experiment demonstrates, unusual juxtapositions of these 'equivalent' forms can be read almost as skillfully as the most regular and common varieties of text.

Summary

The 216 *S*s searched for words in passages of text, some of which were printed in normal upper or lower case, others with alternate letters in upper and lower case or with alternate letters varying in size. The numbers of words identified in a given time for the varying conditions support the experimental hypothesis that the relative size of alternate letters, rather than the alternation of case, accounts for differences between conditions. The results were taken as supporting the view that alternative sets of features for the same letter sequence are treated as functionally equivalent and that readers identify words by (1) discriminating these feature sets and (2) integrating them for identification of the word as a whole.

CUE SELECTION IN PAIRED-ASSOCIATE LEARNING: LENGTH OF STUDY INTERVAL

By Eugene A. Lovelace and Bennet Greenberg,
University of Virginia

Two recent studies indicate that when *S*s learn a list of verbal paired associates in which the stimulus items are poorly integrated, the *S*s often engage in stimulus selection utilizing the initial letter as the functional stimulus.[1] In these two studies the presentation rates were 2 or 1.67 sec. There is the possibility that a selection strategy was markedly enhanced for poorly integrated stimuli by the very limited duration of the study interval. If the *S*s were given additional time during the study interval, performance on the paired-associate (PA) task would clearly improve; the question is whether *S*s would make use of this time to integrate the stimulus or simply to strengthen the association of the functional stimulus to the response term.

In the present study one group of *S*s received three PA acquisition trials on six trigram-digit pairs with 2-sec. study intervals, and a second group received three acquisition trials with 6-sec. study intervals. Following these learning trials each *S* received three special test (ST) trials on which stimulus integration or selection were assessed by showing *S* the individual letters of the trigrams and asking him to provide the missing letters of the trigram and the proper digit response.

Method

Materials and design. The stimuli were six consonantal trigrams of 45% association value in Witmer.[2] The responses were single-digit numbers. No stimulus element (letter) was repeated within the list. The PA learning was by the study-test method using three different orders of presentation. For the three ST trials following PA learning, each stimulus was photographed three times,

* Received for publication December 19, 1968.

[1] L. Postman and R. Greenbloom, Conditions of cue selection in the acquisition of paired-associate lists, *J. exp. Psychol.*, 73, 1967, 91–100; E. A. Lovelace and E. M. Blass, Utilization of stimulus elements in paried-associate learning, *J. exp. Psychol.*, 76, 1968, 596–600.

[2] L. R. Witmer, The association value of three-place consonant syllables, *J. genet. Psychol.*, 1935, 47, 337–360.

once with each letter present and the two missing letters replaced by black squares. Within each ST trial, *Ss* were presented the elements in the first position (Letter 1, at Cue [Position] 1) of two trigrams, those in the second position (Letter 2, at Cue [Position] 2) of two other trigrams, and the third position (Letter 3, at Cue [Position] 3) of the remaining trigrams. Across the three ST trials, each *S* received each element of every stimulus. The order of element postions was randomly determined on ST trials. The order of the ST trials was counterbalanced across *Ss* within each group.

Procedure. For the 18 *Ss* of Condition S, study portions of the PA acquisition trials were presented at a 6-sec. rate; the 18 *Ss* of Condition F received these at a 2-sec. rate. In both conditions the test portions of the PA trials were given at a 2-sec. rate. Items were projected on a translucent screen with a Kodak Carousel slide projector. After the test portion of the third learning trial, the self-paced ST trials were described to the *Ss*.

Subjects. The *Ss* were 36 men who were students in an introductory psychology course. None had served in experiments involving trigram stimuli or PA procedures.

Results

Acquisition. As expected, the variation in study interval resulted in significant differences in performance on the last test trial of acquisition ($t = 3.65$, $df = 34$, $p < .001$). The mean numbers of correct responses were 3.3 and 5.0 for Conditions F and S respectively. Twelve of the 18 *Ss* in Condition S had one or more perfect trial, whereas this occurred for only one *S* in Condition F.

ST trials. Performance, both for digit recall and letter reproduction, was examined for ST Trials 1, 2, and 3. No systematic changes with ST trial were observed for either measure; therefore, all analyses reported hereafter result from combining the data of these three trials.

The mean numbers of additional letters correctly reproduced appear in Table I. The difference between Conditions S and F was not significant ($F = 2.43$, $df = 1/34$, $p >$ .10). Neither cue position

TABLE I

Mean Numbers of Additional Letters Correctly Reproduced

	Letter reproduced	Cue position 1	2	3
Condition F (2-sec. rate)	1		2.11	2.17
	2	2.17		1.61
	3	2.00	1.61	
Condition S (6-sec. rate)	1		2.83	2.78
	2	2.39		2.56
	3	2.61	2.50	

nor the interaction of cue position with study-rate condition was a significant source of variance (both Fs < 1). For Ss in both study-rate conditions the first letter (Letter 1) was more often recalled than the second or third letter (Letters 2 or 3), which were recalled about equally often. These differences were statistically significant for Letter 1 vs. Letter 3 ($t = 2.22$, $df = 35$, $p < .025$) and for Letter 1 vs. Letter 2 ($t = 2.50$, $df = 35$, $p < .01$).

The dashed lines of Fig. 1 show for each condition the percentage of all items on ST trials to which Ss recalled 0, 1, or 2 additional letters as a function of the position of the cue letter presented. The solid lines indicate the percentage of the stimuli at each cue position to which Ss correctly recalled the digit response as a function of the number of additional stimulus elements S was able to report. If reproduction of stimulus elements were not related to digit recall, the two sets of curves in Fig. 1 would have the same form. It is clear that in all cases more digits were correctly recalled when both additional letters were reproduced than would be

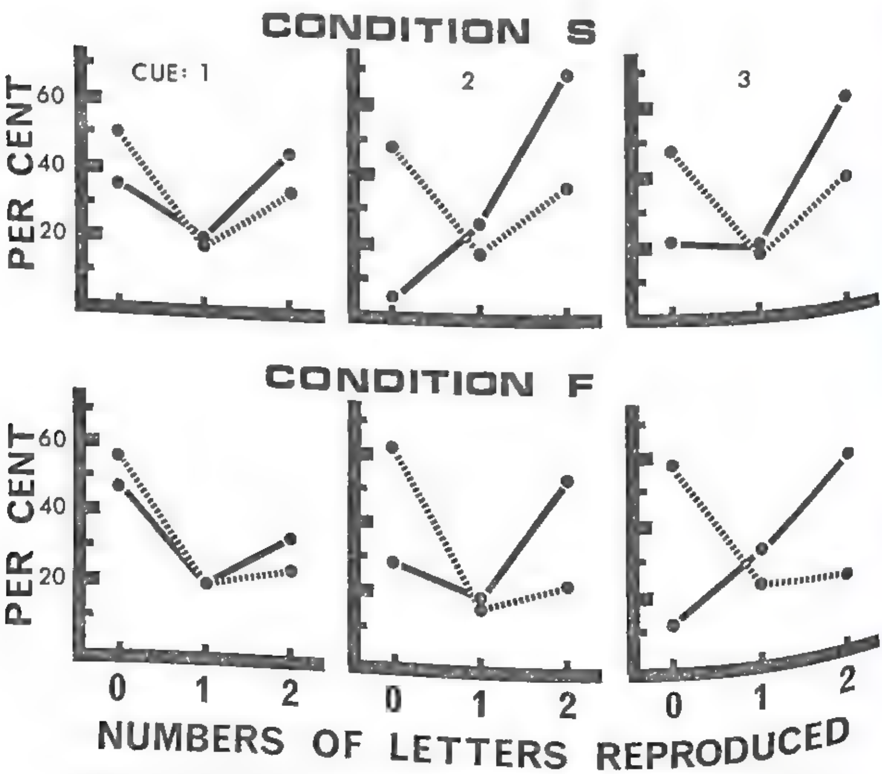


Fig. 1. Distributions of Letter Reproductions and Digit Recalls

expected, whereas fewer digits were recalled when no additional letters were given. These discrepancies in the forms of the curves are greater at Cue Positions 2 and 3 than at Cue Position 1.

Postman and Greenbloom have suggested as a criterion for stimulus selection that the *S* must give the correct digit but be unable to properly reproduce either of the other stimulus elements.[3] This assures that *S* did not mediate the association during the ST trials. Table II displays the number and conditional probabilities of *S*s correctly recalling the digit at each cue position given that 0, 1, or 2 additional letters were reproduced. When *S*s were able to reproduce two additional letters, they were able to give the correct digit about an equal proportion of the time at each cue position. When only one additional letter was reported, there appears to have been a slight trend toward better performance at lower-numbered cue positions. It is clear that when *S*s were unable to reproduce any additional letters, correct responses were not equally probable for the three cue positions. In both groups correct responses occurred the greatest proportion of the time to first-position cues. That is, in both conditions there was evidence of selection of the initial letter as a functional cue.

Discussion

The highly significant differences in correct responses during acquisition and on ST trials (both *p*s < .001) indicate that the *S*s in Condition S learned more about the paired associates than *S*s in Condition F. Despite this greater associative learning, the nonsignificant difference for the two groups on the reproduction of

TABLE II

Numbers and Conditional Probabilities of Digit Recall Given Number of Additional Letters Correctly Reproduced

	Number of additional letters correctly reproduced	Cue position 1		2		3	
		N	*p*	*N*	*p*	*N*	*p*
Condition F (2-sec. rate)	0	60	.37	66	.14	62	.05
	1	21	.43	17	.35	24	.33
	2	27	.59	25	.68	22	.64
Condition S (6-sec. rate)	0	54	.50	51	.04	51	.23
	1	18	.78	18	.72	18	.61
	2	36	.94	39	.87	39	.90

[3] Postman and Greenbloom, *op. cit.*, 92.

stimulus letters indicates that *Ss* in the two groups did not differ very appreciably in what they learned about the stimuli. If such differences in associative learning had been removed by giving more trials to *Ss* in Condition F, these differences in letter reproduction could only be expected to diminish. In neither condition were the *Ss* able to produce a majority of the letters, and the performance of *Ss* in Condition S (44% recall of additional letters) gives further evidence of failure to integrate the stimuli. With a roughly comparable level of PA learning, Postman and Greenbloom report that 83% of the additional letters were recalled when the stimulus materials were such as to produce little stimulus selection, but only 46% when considerable selection of the first letter occurred.[4] Furthermore, in the present study the proportions of the time that *Ss* were able to give the correct digit in response to a cue in the first position when unable to provide any additional letters were not even in the direction which would indicate decreased selection with a slower study rate.

In light of the large and reliable increases in associative learning produced by a reduction of the study rate, it is clear that the slight increase in stimulus integration was not commensurate with this change. It is likely that the slight increase resulted from incidental learning rather than changes in the *Ss*' strategies; to the extent that the stimulus materials of a PA task result in cue selection, they apparently do so even at very slow presentation rates.

Summary

Thirty-six *Ss* received three study-test PA acquisition trials on six trigram-digit pairs; 18 *Ss* received a 2-sec. rate on the study portions and 18 *Ss* received a 6-sec. rate. The *Ss* were then shown individual letters of the trigrams and tested for their ability to provide the digit and to produce the additional letters. The slower rate resulted in significantly more correct responses both in PA acquisition and to the individual letters (both $ps < .001$) but led to only a small, nonsignificant increase in letter reproduction. To the extent that cue selection occurs in a PA learning situation, it apparently occurs even at very slow rates of presentation.

[4] Postman and Greenbloom, *op. cit.*, 93.

ON RECOGNITION AS A DISCRETE EVENT

By Richard L. Taylor, University of Oregon

It is generally possible to accurately describe a current visual stimulus. It is also possible to describe that stimulus after it has been removed from view. Since the two descriptions are likely to differ in detail, if not in accuracy, it is tempting to ask whether the discrepancies arise during recall, and therefore are related to the process of forgetting, or whether they reflect something about how the stimulus was originally encoded. Implicit here is the notion that forgetting and errors in perception are independent. But it is also difficult to imagine how we are able to recognize familiar objects and people without there having been some antecedent interplay between sensory information and the contents of memory. In some sense, when we recognize we have also both remembered and perceived. The difficult question here is that of the conditions under which performance may be affected by the particular joint contents of memory and visual sensory analysis; *i.e.* the question is what conditions combine to produce recognition.

The notion that recognition involves an interaction between sensory and mnemonic messages has been advanced by a number of investigators, including Bartlett, Hoffding, and Woods.[1] More recently, MacKay has made the positive suggestion that incoming sensory information is operated on by a comparator whose function is to organize one's representation of the external world in conjunction with adaptive activity;[2] Sternberg apparently regards the

* Received for publication December 16, 1968. This paper is a partial report of the author's doctoral dissertation submitted to the University of Oregon. The author is indebted to Fred Attneave for sponsoring the work, to Michael Posner for supervising the research prior to his departure for the APRU-MRC in England, to Ray Hyman, for the use (in his absence) of his tachistoscope, to Gerald Reicher for many helpful suggestions, and to Jacob Beck for assistance. The research was supported by the Advanced Research Projects Agency of the Department of Defense, monitored by the Air Force Office of Scientific Research under Contract F44620-67-C-0099. The author is currently Research Associate of the Veterans Administration Hospital, Syracuse, New York.

[1] F. C. Bartlett, *Remembering*, 1932; H. Hoffding, *Outlines of Psychology*, 1891; and E. L. Woods. An experimental analysis of the process of recognizing, this Journal, 26, 1915, 313–327.

[2] D. M. MacKay, Ways of looking at perception, in W. Wathen-Dunn (ed.), *Models for the Perception of Speech and Visual Form*, 1967; D. M. MacKay,

matching of visually coded stimuli with scored test criteria as a basic process;[3] while in auditory theory Halle and Stevens have taken as a key notion that sensory input is compared with internally generated signals.[4]

If such a comparative process is accessible to behavioral observation, then for reasons of biological utility it should be most evident when the contents of memory and visual input are similar. In the present experiments an *S*'s task was to make Same-Different judgments of letter pairs and to respond by pressing one of two levers on each trial. Of interest were those RT trials when the 'Same' lever was pressed correctly since *S* was presumably responding either to (*a*) a stored letter, (*b*) a current visual stimulus, or to (*c*) both a stored and a current visual stimulus which were redundant.

Experiment I

Procedure. An *S* was required to peer into the viewing port of a three-channel tachistoscope, note the occurrence of a probe item, and then decide if any other stimulus appearing in a trial was the same or not. If there was a match, *S* pressed a lever marked "Same." If not, *S* pressed a lever marked "Different." Fig. 1 shows the relative orientation and location of the stimulus events in a typical trial. Note that the probe always appeared in the center 'window,' while the stored and adjacent targets could appear in either the left or right windows during a trial. Presentation times for the item to be stored, for the checkerboard-like interpolated pattern, and for the probe and adjacent items were, respectively, 500 msec., 3 sec., and 2 sec. An *S*'s response and RT were recorded from onset of the probe and adjacent display. There were 144 Same-Different trials in which the 72 Same trials were divided into (*a*) 24 memory matches, (*b*) 24 adjacent matches (as shown in Fig. 1), and (*c*) 24 trials in which both the stored and adjacent items matched the probe. An *S* was exposed only to three letters (B, S, A, or T, N, R) in the session, and each letter appeared equally often in every window. Order of presentation was haphazard.

Subjects. Six students were recruited arbitrarily from among those who were experienced in making Same-Different RT judgments in other tasks. Each was paid $1.50.

Results. The overall average error rate was 3.7%, and those trials were not analyzed. The results for correct responses are shown for each *S* in Table I. The data were consistent for all *S*s tested. Redun-

Cerebral organization and the conscious control of action, in J. Eccles (ed.), *Brain and Conscious Experience*, 1966.

[3] S. Sternberg, Two operations in character recognition, *Percept. & Psychophys.*, 2, 1967, 45-53.

[4] M. Halle, On the basis of phonology, in J. A. Fodor and J. J. Katz (eds.), *The Structure of Language*, 1964; K. N. Stevens, Toward a model for speech recognition, *J. acoust. Soc. Amer.*, 32, 1960, 47-55.

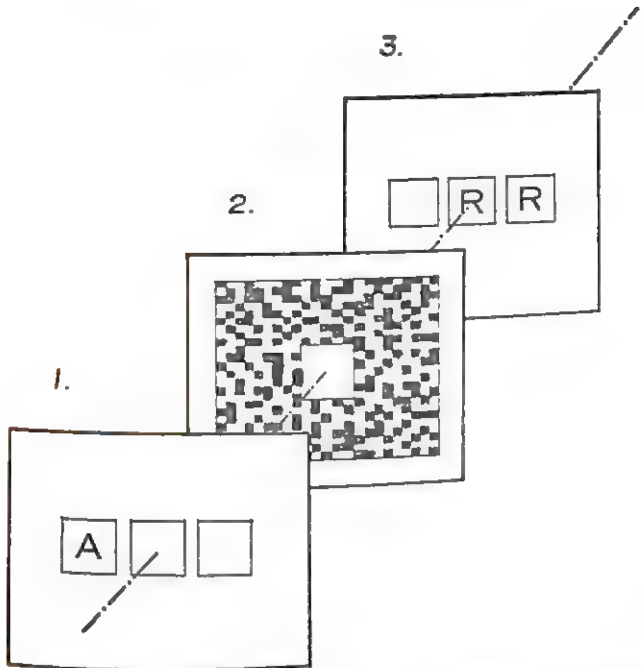


FIG. 1. RELATIVE ORIENTATION AND LOCATION OF STIMULI IN A TYPICAL TRIAL SEQUENCE

dant matching times were faster than memory matches ($t_5 = 4.10$, $p < .01$) and adjacent matches ($t_5 = 18.75$, $p < .001$). The apparent differences between memory and adjacent RTs, and between overall Same and Different RTs, replicate previous findings made under similar conditions.[5]

EXPERIMENT II

It was not possible to determine from Experiment I if identical recalled and perceived stimuli facilitated any response other than a Same judgment. It was also unclear whether the source of the matching item was irrelevant. That is, such an effect might be obtained in any task involving redundant items regardless of whether one item is recalled and the other analyzed from a current visual array. Although there is evidence that identification errors are reduced by redundant visual targets, when an *S* is asked to overtly identify stimuli it is not clear whether the requirement to tap memory in order to come up with a name also enables *S* to benefit from the redundancy. Thus, when an *S* is not required to

[5] *E.g.* R. L. Taylor, A comparison of short term memory and visual sensory analysis as sources of information, *J. exp. Psychol.*, in press.

TABLE I

Median Reaction Times for Correct Responses: Experiment I

Ss	Same responses: Memory	Adjacent	Redundant	Different responses
1	570	620	515	625
2	507	585	492	590
3	525	560	442	587
4	472	565	445	635
5	485	540	405	645
6	490	565	460	612
Means	508	573	460	616

overtly identify stimuli, do redundant simultaneously presented targets facilitate performance? The purpose of Experiment II was to test the reliability of the Experiment I data by using *S*s who were not experienced in making Same-Different RT judgments, to determine if identical adjacent and recalled items also facilitate a Different response, and to make comparisons for a task where three items are presented simultaneously.

Procedure and subjects. Twelve *S*s were recruited for the study by the student employment bureau of the University of Oregon, and each S was paid $1.50. The basic task for six *S*s (Mixed group) was similar to that in Experiment I except that, of the 72 Different trials, 24 were trials where the adjacent and stored item were identical. For the other six *S*s (Simultaneous group), the basic task was similar to that for the Mixed group except that the three stimuli were presented together after the checkerboard pattern. Thus, instead of memory and adjacent matches there were either left or right matches, depending on the window that the matching item appeared in, relative to the probe.

There were, then, five conditions of interest. Imagine three windows in which stimuli can appear: — — —. The probe item (A) always appeared in the center window so that with the Simultaneous group, for example, the matching item could appear in the left window thus: A A B. The other conditions, analogously, were B A A, A A A, B A B, and B A C. For the Mixed group, one need only substitute memory match for left, and adjacent match for right, to employ the same notation.

Results. Median RTs were computed for correct responses in each condition. Five of the six *S*s in the Mixed group had faster A A A RTs than the fastest of their other times, which supports the similar finding in Experiment I. The *S*s were, however, significantly slower in pressing the Different lever when the two comparison stimuli were redundant ($t_5 = 3.13$, $p < .025$). There were no significant differences within Same or Different response conditions for the Simultaneous group. These data are summarized in Table II.

TABLE II

MEAN REACTION TIMES FOR CORRECT RESPONSES: EXPERIMENT II

	Condition				
	Same			Different	
	A A B	B A A	A A A	B A B	B A C
Mixed group	576	628	541	701	662
Simultaneous group	602	584	588	687	659

DISCUSSION

Suppose that the *S*s always test for a match by comparing pairs of stimuli: the probe and stored item first with probability p; the probe and adjacent visual item first with probability $1-p$. Since the outcome of each test could be either successful or not, the distribution of RTs to either type of match alone would be bimodal. On redundant trials, however, the first test made would always be correct, thus producing faster average RTs than either memory or adjacent matches alone. This interpretation would account for (*a*) faster overall memory-match RTs (if $p > .50$), (*b*) faster Same than Different RTs, and (*c*) faster overall RTs under the redundant conditions. But it has been found that memory matches produce faster absolute Same RTs than do adjacent matches when an *S* must consider only one possible type of match under conditions otherwise similar to those in the previous experiments,[6] which suggests that the probability interpretation above is not correct. That is, an *S* is not faster in making a memory match for the reason that his first test is more likely to be a memory-probe test. It therefore seems implausible that *S*s were faster in certain redundant conditions here because of a difference in accuracy of their first test match. A somewhat more subtle basis for evaluating the credibility of the probability hypothesis outlined above rests on the well-established relation between event probability and RT. Any biasing there may be for a particular type of match ought both to produce faster RTs to that type of match and to permit less potential gain from items redundant with another type of match. But in Experiment II the left- and right-window matches pro-

[6] Taylor, *loc. cit.* See also W. G. Chase and M. I. Posner, The effect of auditory and visual confusability on visual and memory search tasks, a paper read at the MPA Convention, 1965; and E. Tulving and P. H. Lindsay, Identification of simultaneously presented simple visual and auditory stimuli, *Acta Psychol.* 27, 1967, 101–109.

duced nearly equivalent RTs (suggesting that for the Simultaneous conditions the value of p was more nearly .50), while there was no overall redundant gain observed.

Suppose, however, that instead of making pair-by-pair Same-Different judgments, the *S*s make an initial consistency test between stored and visual information. Only if the results were inconclusive, or only if the task required more than a consistency test, would the *S*s proceed further. If the two visual items were tested in parallel, then this hypothesis could account for (*a*) faster overall Same RTs, (*b*) faster overall memory-match RTs, (*c*) facilitation of Same RTs by redundant items, (*d*) inhibition of Different RTs, and (*e*) the failure to observe analogous effects in the Simultaneous condition of Experiment II. This notion fits the data rather well, as a first approximation, and if correct would point to the existence of a mechanism which responds selectively to recurrent events. Experiment II indicates that performance gains attributable to redundant items may be a memory-dependent phenomenon. That is, in order to benefit in a skilled-performance task, an *S* must either have an implicit criterion for redundancy stored beforehand or be allowed to take a 'second look' at simultaneously presented stimuli. This is consistent with the notion that under marginal viewing conditions, *S*s benefit from redundant visual targets in an identification task much as if they had as many independent 'looks' as there are targets.[7]

Summary

Two experiments involved Same-Different RT judgments of a probe and target letter. On each trial one target was stored for 3 sec. while another target was shown simultaneously with the probe. An *S* pressed the "Same" lever if either target was the same and the "Different" lever if neither was the same as the probe. Memory matches were faster than adjacent visual matches. When the two targets were redundant, Same RTs were facilitated and Different RTs were inhibited, although these effects were not observed when all three letters were shown simultaneously. The experiments were designed to explore certain implications of the classic hypothesis that stored and sensory information may interact in recognition.

[7] C. W. Eriksen, H. L. Munsinger, and T. S. Greenspon, Identification versus same-different judgments: An interpretation in terms of uncorrelated perceptual error, *J. exp. Psychol.*, 72, 1966, 20–25.

Although the data were consistent with that notion, consideration was also given to the possibility that there is a mechanism which responds selectively to similarities between stored and visual sensory information. Output from such a mechanism could either facilitate or inhibit performance, depending on whether categorization or stimulus discrimination was appropriate to the task.

NOTES AND DISCUSSION

PSYCHOPHYSICAL EVALUATION OF FEEDBACK PHENOMENA AS RELATED TO PRECISION OF FORCE EMISSION: SOME METHODOLOGICAL CONSIDERATIONS

Recognition of the importance of feedback considerations to behavioral theorizing and investigation goes back at least as far as John Dewey, who wrote: ". . . the reflex arc idea leaves us with a disjointed psychology. . . . [In] its failure to see that the arc of which it talks is virtually a circuit, a continual reconstitution, it breaks continuity and leaves us nothing but a series of jerks. . . ."[1] In more recent years, active experimentation bearing upon feedback phenomena has reflected the diverse interests of neuropsychology,[2] engineering psychology,[3] learning theory,[4] and sensory psychology.[5] (The "learning theory" reference, in particular, bears upon relations among cutaneous, kinesthetic, and exteroceptive feedback.) Taken collectively, these studies may be viewed as representing a variety of approaches which have in common the fact that they seek to describe behavior at a level more adequate than the simplistic reflexology so earnestly criticized by Dewey. These feedback studies also have implications for problems in motor performance which can be anticipated in altered environments, such as is characteristic of space.

The primary purpose of this paper was to describe two experiments using psychophysical techniques as a means of assessing

* Received for publication October 9, 1968. The research reported was supported at Princeton University by the Office of Naval Research, Physiological Psychology Branch, under contract NONR 1858(19). The authors acknowledge with appreciation the statistical assistance rendered by Bertram L. Koslin and Richard Pargament. Thanks are also due Carl E. Sherrick, who was kind enough to read a prepublication draft of the manuscript.

[1] John Dewey, The reflex arc concept in psychology, *Psychol. Rev.*, 3, 1896, 357–370.

[2] Richard Held, Plasticity in sensory-motor systems, *Sci. Amer.*, 213, 1965, 84–94.

[3] Alphonse Chapanis, Engineering psychology, *Annu. Rev. Psychol.*, 14, 1963, 285–318.

[4] J. M. Notterman and D. E. Mintz, *Dynamics of Response*, 1965, 131–150.

[5] J. C. R. Licklider, On psychophysiological models, in W. A. Rosenblith (ed.), *Sensory Communication*, 1961, 49–72.

effects upon performance of a simple motor task, consequent upon the interruption of either cutaneous or visual numeric feedback. The former is representative of internal, response-continuous feedback; the latter, of external, response-completed feedback. A secondary purpose was to provide psychophysical data gathered with a squirrel monkey under feedback conditions which afford a modicum of comparability with human performance. For both of the experiments here reported, *S*'s task was to exert pressure with his forefinger upon an essentially isometric disc, doing so in accordance with *E*'s instructions specifying required force magnitudes. During the response, *S* had available to him feedback of cutaneous and kinesthetic origin. Upon termination of the response, visual numeric feedback (continuous to three significant figures) was presented to *S* by means of a digital voltmeter. Thereby, $\Delta F/F$ functions were obtained. After a baseline was established, the effects of reducing cutaneous feedback by means of local anesthesia (Experiment I) and of eliminating visual numeric information (Experiment II) were separately explored. A $\Delta F/F$ function obtained with a squirrel monkey was compared with that of the curve generated under conditions entailing the absence of numeric feedback.

Experiment I

Reduced Cutaneous Feedback

Subjects. The *Ss* were one woman, age 45, and one man, age 33, each having a normal medical and psychological history.

Apparatus. The apparatus was a modification of the analog-computer system described by Notterman and Mintz;[6] instead of a strain gauge, a Sanborn transducer (Model FTA 100-592) served as the manipulandum. The portion of the transducer upon which *S* exerted pressure was 13 mm. in diameter; excursion of the disc was less than 0.4 mm. for forces up to 200 gm. A Hewlett-Packard Digital Recorder (Model 561B) registered peak force upon response termination, its read-out being instantaneously and noiselessly displayed by a Hewlett-Packard Digital Voltmeter (Model 405-AR) in neon figures 16 mm. high and 11 mm. wide. Surmounting the voltmeter was a horizontal array of five small white lamps, each with a numbered label. The lamps were used to signal which of five different force requirements (standard stimuli) was in effect: 25, 50, 100, 150, or 200 gm.

Procedure. The *S* was seated in such manner as to assure a distance (at slightly below eye-level) of 3 ft. between *S* and the display; an armrest was used to position the limb so that the right forefinger was just above the manipulandum. *S* was then instructed to press the disc with forces matching as closely as possible the standard signalled by each lamp, and to do so as often

[6] Notterman and Mintz, *op. cit.*, 9–19.

as possible, consistent with accuracy. Each session began with the 25-gm. requirement and consisted of two ascending and descending series. Standards were successively shifted after 1-min. response periods; there was no time separation between presentation of different standards. A 1-min. rest period separated ascending and descending portions of each series. This regimen was in effect for 10 successive daily sessions. Approximately ½ hr. prior to each of two subsequent sessions, S's right index finger was injected with approximately 1.25 cc. of 1% xylocaine hydrochloride, a dosage routinely used to establish topical anesthesia.

Results. The *SD* of response forces emitted for each force requirement was computed separately for each of the two *S*s. The *SD*s were then divided by the associated force requirements to yield $\Delta F/F$ ratios. Average curves were computed and appear in Fig. 1; they are based upon performance over the last two preinjection (baseline) sessions and the two postinjection sessions, respectively. The baseline curve departs somewhat from data previously reported,[7] the differences probably being due to changes in the physical characteristics of the manipulandum. Analysis of variance showed the postinjection and preinjection curves to be significantly different ($F = 64.04$, $df = 1/9$, $p < .001$),[8] in spite of the reversal at the 200-gm. level. Rate of responding varied from approximately 42 to 37 presses per minute for one *S*, and from 56 to 46 for the other. In both cases, the decline in rate was associated with increased force requirement.

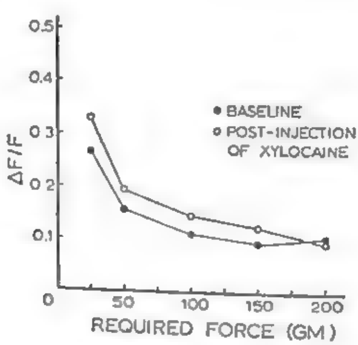


FIG. 1. PRECISION OF FORCE EMISSION AS INFLUENCED BY ANESTHETIC INTERRUPTION OF CUTANEOUS FEEDBACK: EXPERIMENT I

[7] D. E. Mintz and J. M. Notterman, Force differentiation in human subjects, *Psychon. Sci.*, 2, 1965, 289–290.

[8] *F* values are based on pooled estimates of error which include all interactions with *S*s. The reported *F* values are probably conservative, since this procedure tends to underestimate *F*s if nontestable interactions in the denominator are unequal to zero (Q. McNemar, *Psychological Statistics*, 1960, 336).

Discussion. It is evident that deterioration in the precision of force discrimination provided a convenient means for assessing the effects of local anesthesia. Such a result was not unexpected.[9] There are two features of the data, however, which may enhance their importance. (1) The curves are suggestive of means whereby precise evaluation of the specific effects of different local anesthetics, or of different dosages of the same agent, may be made. (2) They are indicative of the manner in which psychophysical procedures may be used to determine the extent of sensory deficit in diseases involving purposive movement (*e.g.* multiple sclerosis) or to trace recovery after appropriate therapy.[10]

Compression of the fingertip flesh and distortion of the skin were observed to be near maximum at 200 gm. The overlap of the two points at this standard thus may be indicative of transition toward reliance upon kinesthetic rather than cutaneous cues, even in the absence of local anesthesia. Such an interpretation is, at present, speculative.

EXPERIMENT II

REMOVAL OF VISUAL NUMERIC FEEDBACK

The same *Ss* and apparatus as in Experiment I were used, with the exception that the digital voltmeter was removed. Upon presentation of the signal, *S* was required to press within a specific range of forces which extended above and below the indicated target force. This range was ± 1 *SD* for the respective targets as computed from the responses of each *S* in the preanesthetic phase of Experiment I. If *S* pressed within the designated range appropriate to each standard, a distinctly audible click instantaneously followed response termination. Otherwise, the end of a response received no notice. Ascending and descending series, response periods, rest periods, and number of daily sessions were the same as for Experiment I.

Results. Weber ratios were computed on the basis of the *SD* of forces actually emitted, as averaged over the last two sessions, and comprise the middle curve in Fig. 2. This curve is significantly different ($F = 13.95$, $df = 1/9$, $p < .01$)[11] from the baseline (lower) curve obtained under conditions of visual numeric feed-

[9] Theodore Marton, Peripheral afferent cutaneous information and the control of voluntary motor activity, unpublished Ph.D. dissertation, Princeton University, 1962. Also, K. A. Provins, Sensory factors in the voluntary application of pressure, *Quart. J. exp. Psychol.*, 9, 1957, 28–41.

[10] D. N. Robinson and A. J. Berman, Some quantitative features of involuntary movement, Final Report, May 1, 1968, Division of Vocational Rehabilitation, Department of Health, Education and Welfare.

[11] See n. 8.

back (Experiment I). Again, rate of responding tended to decline with increased force requirement.

Discussion. Under the particular conditions of this experiment, the effect of removing visual numeric feedback and substituting 'yes-no' auditory information was to increase the $\Delta F/F$ values by approximately 20%. What is perhaps of equal importance in the data is the fact that they permit rough comparison of human performance with that of a monkey's, the lower primate not only routinely doing without the benefit of numeric information but also often serving as the subject-of-choice in the pharmaceutical, neurological, and space research referred to previously.

Accordingly, one male squirrel monkey (about 2½ years old, 475-gm. body weight) was trained to press a disc-shaped manipulandum (19 mm. in diameter) located about 6 in. outside his wire cage. Forces greater than 10 gm. were required in order to obtain 75-mg. banana-flavored food pellets, routine 22-hr. hunger rhythm having been established more than 10 days prior to the onset of conditioning. Following 30 sessions of this procedure, the animal was shifted to band reinforcement, such that reinforcement was delivered only if response force fell within the upper and lower limits specifying the band.[12] Four such bands of required force were investigated: 30–40, 50–60, 70–80, and 90–100 gm. The monkey was exposed to each band until performance stabilized, whereupon he was shifted to the next higher band. The number of daily sessions at each of these bands varied from 14 to 26. All sessions were terminated following delivery of 75 reinforcements or an hour's time, whichever came first. Data

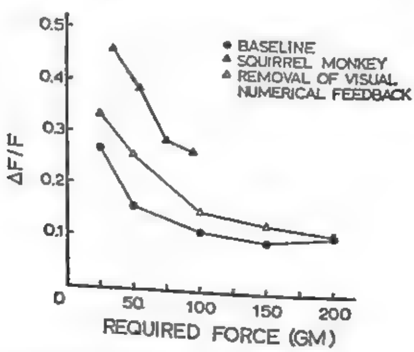


FIG. 2. PRECISION OF FORCE EMISSION AS INFLUENCED BY REMOVAL OF VISUAL NUMERIC FEEDBACK: EXPERIMENT II

[12] Notterman and Mintz, *op. cit.*, 101–110.

analysis was based upon performance averaged over the last five days at each band. The resulting ΔF/F function appears as the upper curve in Fig. 2, the midpoint of each band serving as the standard *F*.

On the average, the monkey's ΔF/F function was about 50% higher than that generated by 'yes-no' feedback data obtained with human *S*s. As is characteristic of animal psychophysics in general, the differences in motivational and reinforcing parameters were such as to render direct comparison with human data quite difficult. Moreover, since the monkey had no armrest, the weight of the extended forelimb could quite readily have influenced kinesthetic feedback and, thereby, the threshold. (The weight factor has special relevance for space research, in that further degradation of performance can probably be anticipated in a gravity-free environment. Such seems reasonable in that the weight of the limb ceases to be of any consequence in the production of reactive forces, thereby probably leading to concomitant loss of kinesthetic information.) It is also true that the superiority of human discrimination in this elementary motor task is in keeping with evolutionary and anthropological suppositions placing emphasis upon manual dexterity in attempting to account for man's development.

Summary

Two experiments using psychophysical procedures to assess the effects of changes in feedback were described. Experiment I was concerned with cutaneous (or response-continuous) information; Experiment II, with visual numeric (or response-completed) information. The two types of feedback have approximately equal influences under the specific conditions reported. In connection with Experiment II, a ΔF/F function, describing Weber ratios for force emission, was obtained with a squirrel monkey and was compared with that obtained for the human *S*s. The possible role of kinesthetic feedback and its relation to weight of the limb was discussed.

University of Pennsylvania
Princeton University

Ross D. L. Filion
Stephen C. Fowler
J. M. Notterman

TIME TO LOCATE PROBE ITEMS IN SHORT LISTS OF DIGITS

Sternberg proposed a theory of the process of retrieval in serial lists—a theory which relates response latency to the length of a list and to the serial position of the item to be located.[1] In the first of his studies cited, the time to indicate whether or not a probe item was present in a previously learned list was shown to increase as a linear function of the length of the list. Since latencies for positive and negative responses did not differ greatly, it was argued that scanning through the series, in both cases, was exhaustive. In the later paper cited, the relationship between probe position and latency was examined directly. The *S*s were asked to name the items following a probe item in a list. There was a pure primacy effect: response latency increased as a linear function of the serial position of probe items. There are two unsatisfactory features of this second experiment, however. (1) There is obviously a danger in seeing the task of indicating presence or absence as equivalent to that of indicating the next item. Using grammatical material, it has been demonstrated that the task of scanning to locate (*i.e.* to indicate presence or absence) may be a component part of the task of scanning to retrieve another item,[2] and the relationship between serial position and search time for the two tasks may thus not be identical. (2) Sternberg's study confounded errors with list length, which may explain why Kennedy failed to reproduce the apparently linear relationship between reaction time and probe position.[3]

In another study, Morin, Derosa, and Stultz examined response latency as a function of probe position, using sets of four digits selected from the numerals 1–8.[4] Closed sets (those in which the digits could be permuted to form a consecutive series) and open sets (those in which the digits would not form a consecutive series), in either increasing, decreasing, or irregular sequences, were used. *S*s indicated presence or absence of test digits by saying either

* Received for publication December 2, 1968.

[1] S. Sternberg, High speed scanning in human memory, *Science*, 153, 1966, 652–654; S. Sternberg, Retrieval of contextual information from memory, *Psychon. Sci.*, 8, 1967, 55–56.

[2] R. A. Kennedy and A. L. Wilkes, An analysis of storage and retrieval processes in memorizing simple sentences, *J. exp. Psychol.*, in press.

[3] R. A. Kennedy, Response latency in the serial learning of short lists, *Brit. J. Psychol.*, 59, 1968, 1–5.

[4] R. E. Morin, D. V. Derosa, and V. Stultz, Recognition, memory, and reaction time, *Acta Psychol.*, 27, 1967, 298–305.

"yes" or "no." The relationship between the serial position of probes and response latency was similar for all subconditions and showed a marked recency effect. In all cases, latency at Position 4 was very much more rapid than at any other position. For open sets, latency at Position 2 was always longest. For closed sets, latencies at Positions 1, 2, and 3 did not differ.

While these findings are obviously in conflict with Sternberg's theory, their generality is somewhat limited, and the present paper set out to examine the relationship between response latency and serial position in lists of varying lengths.

Method

Subjects. The *Ss* were 10 students in the University of Dundee, 4 women and 6 men.

Procedure. Ninety lists of from one to five digits were drawn from random-number tables. There were 30 five-digit lists, 24 four-digit lists, 18 three-digit, 12 two-digit, and 6 one-digit lists. In this way, equal representation of each serial position was maintained. The lists were placed in two random series of 45 lists each. In one series, each list was followed by a probe item present in the list; in the other series, probes were not represented in the lists. The total set of lists, each followed by a probe item, was then randomized and recorded on tape at a rate of one item per sec. Probes followed 2 sec. after a list and were preceded by a warning tone. There was a 4-sec. interval between lists. The *Ss* responded by saying "yes" or "no" to a probe item. The entire testing session was recorded and later processed by playing it through a voice key, where a probe item started a timer and *S*'s response stopped it.

Results

Mean response latency at each serial position for positive probes is shown in Fig. 1. Clearly, for lists containing three, four or five items, the relationship between position and latency was not linear. Three independent analyses of variance were carried out on these lists, factors being subjects and position. In all lists, the effect of position was significant: three-digit, $F\ (2,18) = 4.80$, $p < .05$; four-digit, $F\ (3,27) = 5.50$, $p < .01$; five-digit, $F\ (4,36) = 5.02$, $p < .01$. Between-means comparisons showed that in each case the final position was significantly less than Position 2. Errors were very infrequent ($< 2\%$) and were replaced in the analyses by *S*'s mean score in the relevant condition.

Fig. 2 shows the mean latency for each list length, comparing items to indicate presence and absence. Wilcoxon tests at each length showed the differences between positive and negative probes to be significant for two-digit lists ($p < .01$) and three-digit lists

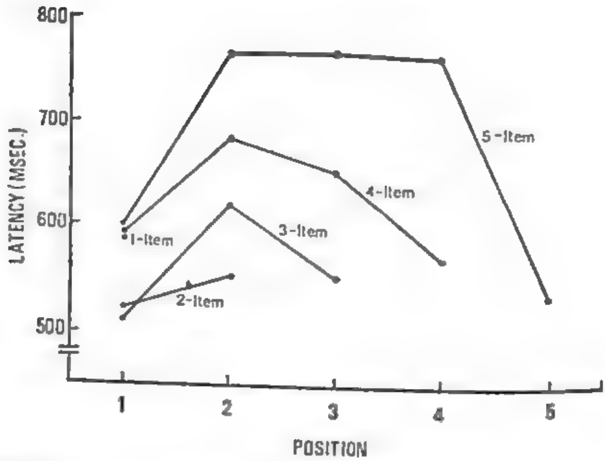


FIG. 1. MEAN RESPONSE LATENCY TO INDICATE PRESENCE OF A PROBE ITEM IN LISTS 1–5 DIGITS LONG AS A FUNCTION OF POSITION OF PROBE WITHIN A LIST

($p < .05$). The difference for one-digit lists also approached significance.

DISCUSSION

The results of the present study were not consistent with predictions made from Sternberg's model. The exhaustive-scanning hypothesis must be rejected on the grounds that times to positive and negative probes for a single list length were not always identical. The serial-comparison aspects of the model were *not* supported by the evidence of a curvilinear relationship between latency and probe position. Rather, the results supported earlier work, using paired associates, which demonstrated that for unstructured lists

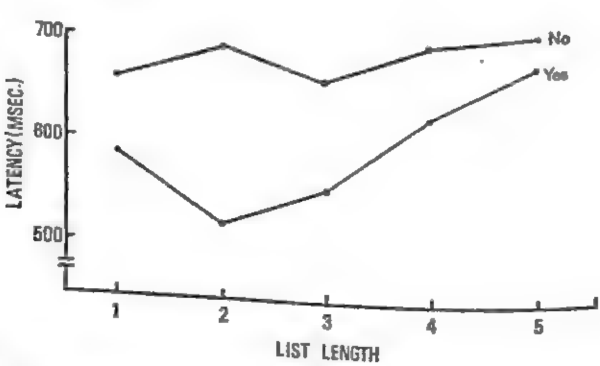


FIG. 2. MEAN RESPONSE LATENCY TO INDICATE PRESENCE ('YES') OR ABSENCE ('NO') OF PROBE ITEMS AS A FUNCTION OF LIST LENGTH

the end positions were more accessible.[5] The apparent differences in latency at Position 1 for the five lists were difficult to interpret but may reflect no more than the rarity of short lists within the total series.

A model incorporating some mechanism providing for priority of access to particular positions within a list is called for. If the material being scanned is structured in some way, as with sentences, then there is evidence that priority of access is given to positions marking grammatical boundaries.[6] In the case of unstructured lists, on the other hand, *S* may impose structure through recoding or by imposing rhythm in rehearsal.[7] The short lists used in the present study were clearly structured by means of two defining points (the beginning and the end), and these points had priority of access. It may well be that for this task the notion of serial scanning (*i.e.* position-by-position matching) does not invariably represent the kind of search process which *S* initiates. Little is known in detail of the process of search in short-term storage, but it is likely to be sensitive not only to the particular decay characteristics of the storage system but also to the mechanisms of rehearsal and re-coding under *S*'s control.[8] The present results suggest that for one type of (relatively fast) auditory input *S* merely categorizes stimuli into two levels of accessibility. Within a given level, search time revealed no obvious scanning strategy.

Summary

Response latency was measured as *Ss* indicated presence or absence of single items after learning short lists of digits (1–5 items). The relationship between latency and the serial position of probe items was not linear: items from the beginning and end were retrieved rapidly, and no systematic relationship appeared for other positions. In general, *Ss* were faster to indicate presence than absence of a probe item, particularly for short lists.

University of Dundee, Scotland

R. A. Kennedy
D. Hamilton

[5] R. A. Kennedy and A. Galloway, Retrieval time and the effect of serial position, this Journal, 80, 1967, 430–433.

[6] A. L. Wilkes and R. A. Kennedy, The relationship between pausing and retrieval latency in sentences of varying grammatical form, *J. exp. Psychol.*, in press.

[7] U. Neisser, *Cognitive Psychology*, 1966.

[8] R. C. Atkinson and R. M. Shiffrin, Human memory: A proposed system, in K. W. Spence and J. T. Spence (eds.), *The Psychology of Learning and Motivation*, 1968.

RETROACTIVE INHIBITION, CUE SELECTION, AND DEGREE OF LEARNING

Schneider and Houston have demonstrated that in a paired-associate situation, stimulus selection during List-2 learning can influence amount of obtained retroactive inhibition (RI).[1] Their experiment was based on this analysis:

> . . . Assume different responses are learned in two successive trigram-adjective lists. In addition, assume that the second-list trigrams are surrounded by colors. If *S* selects a component of the second-list stimulus which is identical to the stimulus utilized in the learning of the first list (viz., the trigrams) then the situation corresponds to the A-B, A-C paradigm. Considerable unlearning and competition are expected to occur in this situation. On the other hand, if *S* utilizes a component of the second-list stimulus which is different from that used in first-list learning (viz., the colors) then the situation corresponds to the A-B, C-D paradigm and much less unlearning is expected. Partial attention to both components might result in intermediate amounts of RI.[2]

As predicted, their data indicated that RI increased as pressure to attend to the trigrams during List-2 learning was increased.

The purpose of the present experiment was twofold. (1) It attempted to replicate the findings of the earlier study. (2) It varied the degree of List-2 learning. The *Ss* learned List 2 either to a criterion of one perfect trial or to a criterion of one perfect trial plus 10 additional trials. The question at issue was whether the overlearning condition might produce an increase in RI in some of the pressure conditions. Specifically, James and Greeno have found that during overlearning, *Ss* will utilize, or begin to attend to, cues which were ignored prior to the attainment of the criterion.[3] Thus, in the present study, one might expect overlearning to produce a shift in attention to the trigrams and a corresponding increase in RI in those conditions where pressure to attend to the colors exists.

Method

Design, lists, and procedure. The experiment involved a 2 × 6 design with one factor referring to the variation in degree of List-2 learning (Criterion

* Received for publication September 19, 1968. The work reported was supported by a Public Health Service grant (15336-01) to the second author.

[1] N. G. Schneider and J. P. Houston, Stimulus selection and retroactive inhibition, *J. exp. Psychol.*, 77, 1968, 166–167.

[2] Schneider and Houston, *op. cit.*, 166.

[3] C. T. James and J. G. Greeno, Stimulus selection of different stages of paired-associate learning, *J. exp. Psychol.*, 74, 1967, 75–83.

and Overlearning conditions). The other factor referred to six different degrees of pressure to attend to the trigrams during List-2 learning. All *Ss* learned two successive paired-associate lists and were then tested immediately for List-1 recall. The trigram-adjective lists (1) were identical in all 12 conditions. The List-2 responses in all conditions were adjectives which differed from the List-1 responses. In the two A groups (Criterion and Overlearning), the List-2 stimuli were colors. These groups corresponded to the A-B, C-D paradigm. The List-2 stimuli in the B–E groups were the List-1 trigrams surrounded by different colors. *Ss* in these groups were given different instructions on List-2 learning. *Ss* in the B groups were instructed to attend to the colors and to ignore the trigrams during List-2 learning. They were informed that they would be tested with the colors alone following the completion of learning. In the C groups *Ss* were given no special instructions on List-2 learning. In the D groups *Ss* were asked to attend to both the colors and the trigrams and were informed that they would be tested with each type of cue following learning. The E groups were instructed to attend to the trigrams alone and were told that these units would serve as the cues in the ensuing test. The trigrams alone, without the colors, served as the List-2 stimuli in the F groups, which corresponded to the traditional A-B, A-C interference paradigm. The assumption behind the development of these six conditions, as detailed in the Schneider and Houston study,[4] was that pressure to attend to the trigram during List-2 learning would increase as one moved from the A to the F groups. Accordingly, the predictions in the present study were as follows. When List-2 learning is taken to a criterion of one perfect trial, a decrease in List-1 recall was predicted as one moves from the A to the F groups. This half of the experiment, with one or two minor changes, replicated the Schneider and Houston study.[5] Because of the mechanism suggested by James and Greeno,[6] whereby *Ss* begin to attend to previously unselected stimulus components during overlearning, it was predicted that overlearning in the present study would lead, relative to the Criterion conditions, to a decrease in List-1 recall in those conditions where the trigrams were available and pressure to attend to the colors existed during List-2 learning, *i.e.* in the B and C groups.

Subjects and procedure. The *Ss* were 120 undergraduates at the University of California, Santa Barbara. The 12 conditions were randomized so that they formed 10 blocks of conditions. Each condition occurred once in each block. *Ss* were assigned to this arrangement as they appeared in the laboratory. The lists used in this study were those employed and described by Barnes and Underwood.[7] They consisted of one set of eight trigrams, two sets of eight different adjectives, and eight different colors. Five presentation orders of each list were developed, and each was used equally frequently as the starting order. All learning was at a 2:2-sec. presentation rate. The intertrial and interlist intervals were 4 sec. and 1 min. respectively. List-1 learning was taken to a criterion of one perfect trial. List-2 learning was taken either to this criterion or to this criterion plus 10 overlearning trials. *Ss* were presented the

[4] Schneider and Houston, *op. cit.*, 167.
[5] Schneider and Houston, *op. cit.*, 166–167.
[6] James and Greeno, *op. cit.*, 75–83.
[7] J. M. Barnes and B. J. Underwood, "Fate" of first-list associations in transfer theory, *J. exp. Psychol.*, 58, 1959, 97–105.

trigram stimuli 1 min. after they completed their List-2 learning, and they were asked to recall the correct List-1 responses. Each trigram was presented for 2 sec. The responses were not exposed during the recall test. Between the presentation of succeeding stimuli there were 2-sec. rest intervals.

RESULTS AND DISCUSSION

As expected, an analysis of variance revealed no significant differences among the 12 List-1 trials to criterion values: $F = 2.04$. Table I contains the List-2 trials to criterion values and the mean List-1 recall values. An analysis of variance applied to the List-2 learning data indicated a significant group effect: $F\ (5,108) = 9.58$, $p < .01$. This was probably due to the difference between the F and A groups, since the F group corresponded to the highly negative A-B, A-C transfer paradigm while the A group was similar to the traditional A-B, C-D control transfer paradigm. When the F and A groups were excluded from the analysis, significance was not attained: $F\ (3,72) = 2.40$. This suggests that the instructional pressure involved in groups B–E did not, in fact, affect List-2 learning. The degree of learning and interaction effects were not significant: $F\text{s} < 3.04$.

The List-1 recall data replicated the Schneider and Houston finding.[8] That is, there appears to have been a decrease in List-1 recall from Group A to F: $F(5,108) = 17.83$, $p < .01$. Thus the basic hypothesis, which states that memory losses may be strongly affected by cue selection occurring during interpolated learning, receives additional support from the present study. Contrary to expectations, however, the degree of learning and the interaction effects were not significant: $F\text{s} < 1.76$. The expected increases in RI did not appear. There are a number of possible explanations for this fact. For example, it could be that a shift to the trigrams during overlearning did not occur in the B and C groups. The *S*s may have avoided the trigrams in this

TABLE I

MEAN LIST-2 LEARNING AND LIST-1 RECALL VALUES

	Group					
	A	B	C	D	E	F
List-2 trials to criterion						
Criterion	5.90	9.40	11.30	6.80	11.40	12.40
Overlearning	4.30	7.30	8.90	7.30	8.80	13.20
Correct List-1 recall						
Criterion	5.60	5.70	3.60	4.10	3.20	2.60
Overlearning	5.90	5.80	5.00	2.90	2.80	1.90

[8] Schneider and Houston, *op. cit.*, 167.

experiment, even during overlearning, because to attend to them would have established a competing negative transfer relationship between List 1 and List 2. On the other hand, one might argue that the shift to previously ignored stimulus components may not be a reliable or general phenomenon. Houston, in fact, reports data which fail to display such a shift.[9] Further research should be directed toward the resolution of these ambiguities.

Summary

Subjects learned two successive paired-associate lists and were then tested for recall of the List-1 responses. It was found that retroactive inhibition increased as pressure was increased to attend to the List-2 stimulus components identical to those utilized during List-1 learning. This result supports the hypothesis that cue selection during interpolated learning can be an important determinant of retroactive memory losses. Degree of List-2 learning was found to be unrelated to the strength of these effects.

University of California, Los Angeles

Nina G. Schneider
John P. Houston

[9] J. P. Houston, Stimulus selection as influenced by degrees of learning, attention, prior associations, and experience with the stimulus components, *J. exp. Psychol.*, 75, 1967, 509–516.

AVERSIVENESS OF NONREWARD CUES AS A FUNCTION OF DEGREE OF EXPOSURE

The concept of latent extinction is derived from experiments which follow a specific procedure: after acquisition, animals in the experimental group are placed directly into a goal box without reward for a number of trials prior to initiation of the extinction trials. Animals treated in this manner show more rapid extinction during the test trials than those not given previous nonrewarded placements. Among the many interpretations of this latent-extinction phenomenon is one which focuses on the aversive effects of frustrative nonreward induced by the nonrewarded placements.[1]

* Received for publication 28 August 1968. The research reported was supported by a grant from the National Research Council of Canada, APA-247.

[1] A. Amsel, Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension, *Psychol. Rev.*, 69, 1962, 306–328.

Thus, according to the frustration hypothesis, the experiencing of nonreward leads to an aversive drive state known as frustration: the animal, seeking to avoid this aversive state during the extinction period, acquires alternative competing responses which hasten the extinction of the 'correct' response.

Studies which are generally accepted as relevant to the frustration hypothesis demonstrate that the number as well as the duration of preextinction trials contribute to the latent-extinction effect.[2] Such experiments have not considered direct measures of the aversive properties of the nonreward cues. The present experiment was designed to determine if rats are able to associate distinctive cues with different degrees of nonreward and, as a consequence, show differential avoidance of such cues in a free-choice test. More specifically, the aim of this experiment was to determine if the aversiveness of nonreward cues is related to the number of nonrewarded placement trials.

Method

Subjects. The *Ss* consisted of 14 male and 14 female experimentally naïve rats of the Wistar strain. They were about 60 days old at the beginning of the experiment. The males weighed about 250 gm. and the females about 200 gm. at that time.

Apparatus. The test apparatus consisted of a short-stem Y maze, the arms of which were detachable end boxes. The stem was 10 in. long and 4¾ in. wide, and the end boxes were 11 in. × 3 in. All parts of the maze were 6 in. deep and were covered with removable hardware-cloth tops. The stem was painted gray, one of the end boxes was white and the other black. A third box was painted with black and white vertical stripes which were about ¾ in. wide. Time measures were recorded with two hand-operated .01-sec. Lafayette electric timers. A Gralab Universal Timer (Model 165) was used to measure 1-min. intervals during the experiment.

Procedure. All the rats were placed on a daily food ration of 10 gm. four days before the beginning of the test trials. Two days before they were trained and tested, they were given prefeeding and prehandling. They were then randomly assigned to two groups of equal size, with the restriction that the groups be balanced for number of rats of each sex. Each training session consisted of placing each rat six times into the striped box, which always contained a 45-mg. Noyes pellet, five times into the second box, without any reinforcement, and once into the third box, also without any reward. One group received its five nonrewarded placements in the black box; the other group received these five placements in the white box. The order of the 12 placements was varied from day to day. After rat received its 12 placements, it was momentarily returned to the carrying cage while the black box and the

[2] J. A. Dyal, Latent extinction as a function of the number and duration of the pre-extinction exposures, *J. exp. Psychol.*, 63, 1962, 98–104.

white box were placed alongside the stem to form the Y maze. The rat was then placed into the stem so that it faced the end opposite to the end boxes, and was free to wander about the apparatus for 1 min., during which period the time it spent in either end box was recorded on a cumulative timer. After the end of the 1-min. period, the rat was removed and returned to its home cage. This training and testing procedure was continued for 24 days.

Results and Discussion

An analysis of the time spent by the rats in either the more-trained or the less-trained nonrewarded end box revealed that they preferred the less-trained ($F = 8.031$, $df = 1/26$, $p < .01$). An inspection of Fig. 1, which plots the proportion of time spent by the rats in the less-trained nonrewarded end box, by trial blocks, indicates that this preference reached asymptote by the end of the fourth day. There also did not seem to be any decrease of this preference over further trial blocks.

An analysis of the total time which the rats spent in either of the end boxes showed that there was a significant decrease over trials ($F = 14.835$, $df = 5/130$, $p < .01$). With increasing test trials, the rats spent less time entering the end boxes and more time returning to the stem or staying there. It should be noted that the animals spent significantly more time in the black box than in the white box ($F = 37.750$, $df = 1/26$, $p < .001$). Their preference was independent of the number of nonrewarded placements which they received in either box.

The results of this experiment are interesting because they indicate that the animals found the cues associated with five daily

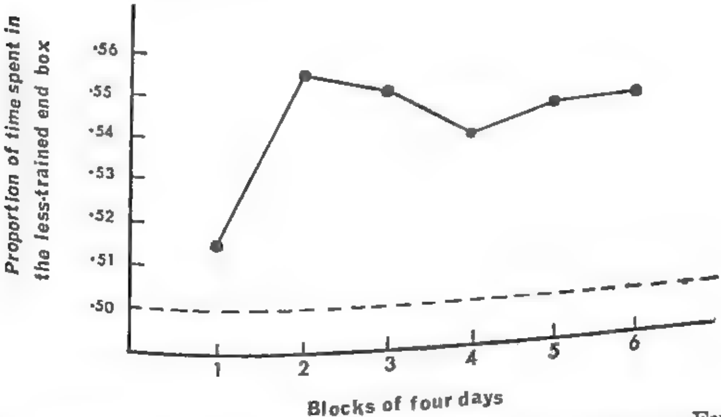


Fig. 1. Proportion of Time Spent in Less-Trained Nonrewarded End Box over Trial Blocks

nonrewarded placements more aversive than those associated with one daily nonrewarded placement. In contrast to the Type II latent-extinction design where an animal is given a choice between entering an end box previously paired with feeding and another box previously paired with nonreward,[3] the present experiment was designed so that the animal was given a choice between two end boxes both of which were associated with nonreward. The other major procedural difference between the present experiment and the latent-extinction studies is that a temporal rather than a frequency-of-choice measure of preference was used in this study.

It is apparent that the animals were able to associate (latently) distinctive cues with various degrees of nonreward and thus exhibited differential avoidance of such cues during a free-choice test. Not only did the rats show a greater preference for the less-trained nonrewarded end box over the more-trained one, but they also spent less time entering either of these end boxes with increasing test trials. More substantively, the aversiveness of the less-trained nonrewarded cues appears to be positively related to the relative number of nonrewarded placements.

Windham College, Putney, Vermont — RODERICK WONG
University of British Columbia — JOHN L. JAMIESON

[3] H. Moltz, Latent extinction and the fractional anticipatory response mechanism, *Psychol. Rev.*, 64, 1957, 229–241.

BOOKS RECEIVED

(The books listed here have not as yet been noted in our pages. Listing here does not, however, preclude their later review.)

Aaronfreed, Justin. *Conduct and Conscience*. New York: Academic Press, 1968. Pp. viii, 405. $12.50.

Bangs, Tina E. *Language and Learning Disorders of the Pre-Academic Child*. New York: Appleton-Century-Crofts, 1968. Pp. 408. $6.50.

Barcroft, T. A. *Topics in Intermediate Statistical Methods*. Ames: Iowa State University Press, 1968. Pp. xii, 129. $5.59.

Barker, Roger G. *Ecological Psychology*. Stanford, Calif.: Stanford University Press, 1968. Pp. vi, 242. $7.50.

Barnes, Elizabeth, ed. *Psychosocial Nursing*. New York: Tavistock, 1968. Pp. xii, 316. Cloth, $10.00; paperback, $4.75.

Barnouw, Erik. *The Golden Web: A History of Broadcasting in the United States, 1933–1953*. New York: Oxford University Press, 1968. Pp. 391. $9.00.

Bayley, David H., and Mendelsohn, Harold. *Minorities and the Police*. New York: The Free Press, 1969. Pp. 209. $8.95.

Beck, Helen L. *Social Services to the Mentally Retarded*. Springfield, Ill.: Charles C Thomas, 1969. Pp. xvi, 207. $8.75.

Bergman, Rita E., ed. *The Sociopath: Selections in Anti-Social Behavior*. New York: Exposition Press, 1968. Pp. 153. $5.00.

Berrien, F. Kenneth. *General and Social Systems*. New Brunswick: Rutgers University Press, 1968. Pp. vii, 231. $9.00.

Bettelheim, Bruno. *The Children of the Dream*. New York: Macmillan, 1969. Pp. xi, 363. $6.95.

Blanck, Rubin, and Blanck, Gertrude. *Marriage and Personal Development*. New York: Columbia University Press, 1968. Pp. xiv, 191. $6.00.

Blood, Robert O., Jr. *Marriage*. (2nd ed.) New York: The Free Press, 1969. Pp. 535. $8.95.

Boe, Erling E., and Church, Russell M. *Punishment: Issues and Experiments*. New York: Appleton-Century-Crofts, 1968. Pp. ix, 329. $4.50.

Borden, G. A., Gregg, R. B., Grove, T. G. *Speech Behavior and Human Interaction*. Englewood Cliffs, N. J.: Prentice-Hall, 1969. Pp. vii, 260. $5.95.

Botwinick, Jack. *Cognitive Processes in Maturity*. New York: Springer, 1967. Pp. x, 212. $5.75.

Brautigam, Walter. *Reaktionen Neurosen Psychopathien: Ein Grundrif der kleinen Psychiatrie*. Germany: Georg Thieme Verlag, 1968. Pp. vii, 175. Paperback.

Briere, Eugene John. *A Psycholinguistic Study of Phonological Interference*. Paris: Mouton, 1968. Pp. 84.

Brodey, Warren M. *Changing the Family*. New York: Clarkson N. Potter, 1968. Pp. 160. $6.00.

Bucknill, John Charles, and Tuke, Daniel H. *A Manual of Psychological Medicine.* (Facsimile of 1858 ed.) New York: Hafner, 1968. Pp. 536. $14.50.

Bühler, Charlotte, and Massarik, Fred, eds. *The Course of Human Life.* New York: Springer, 1968. Pp. v, 422.

Butcher, H. J. *Human Intelligence.* New York: Barnes & Noble, 1968. Pp. 342. $8.75.

Calia, Vincent F., and Wall, Bartholomew D., eds. *Pupil Personnel Administration: New Perspectives and Foundations.* Springfield, Ill.: Charles C Thomas, 1968. Pp. xiv, 329. $11.50.

Catania, A. Charles. *Contemporary Research in Operant Behavior.* Glenview, Ill.: Scott, Foresman and Company, 1968. Pp. 358.

Chassan, J. B. *Research Design in Clinical Psychology and Psychiatry.* New York: Appleton-Century-Crofts, 1967. Pp. xvii, 279. $5.75.

Cherry, Colin. *On Human Communication.* Cambridge: The M.I.T. Press, 1968. Pp. xiv, 337. $2.95.

Chess, Stella, and Thomas, Alexander. *Annual Progress in Child Psychiatry and Child Development.* New York: Bruner, 1968. Pp. x, 565. $15.00.

Cohen, John. *Psychological Time in Health and Disease.* Springfield, Ill.: Charles C Thomas, 1967. Pp. xv, 103. $6.50.

Coopersmith, Stanley. *The Antecedents of Self-esteem.* San Francisco: W. H. Freeman, 1967. Pp. ix, 283. $6.00.

Costigan, Giovanni. *Sigmund Freud.* New York: Collier, 1968. Pp. xiv, 306. $2.45.

Cox, David. *Modern Psychology: The Teachings of Carl Gustav Jung.* New York: Barnes & Noble, 1968. Pp. x, 181. Paperback, $1.25.

Crosby, R. M. N. *The Waysiders.* New York: Delacorte Press, 1968. Pp. 241. $6.00.

Dalton, G. W., Barnes, L. B., and Zaleznik, A. *The Distribution of Authority in Formal Organizations.* Boston: Harvard University Press, 1968. Pp. xii, 229. $6.00.

Despert, J. Louise. *Schizophrenia in Children.* New York: Robert Brunner, 1968. Pp. 215. $6.95.

Dominian, J. *Marital Breakdown.* Baltimore: Penguin Books, 1968. Pp. vii, 172.

Dubin, Robert. *Theory Building.* New York: The Free Press, 1969. Pp. ix, 298. $7.95.

Eisenbud, Jule. *The World of Ted Serios.* New York: Pocket Books, 1968. Pp. xi, 367. $.95.

Esper, Erwin A. *Mentalism and Objectivism in Linguistics.* Vol. 1. New York: American Elsevier, 1968. Pp. 246. $12.50.

Etzkorn, K. Peter. *Georg Simmel: The Conflict in Modern Culture and Other Essays.* New York: Teachers College Press, 1968. Pp. ix, 140. $5.00.

Evans, Richard I. *B. F. Skinner: The Man and His Ideas.* New York: E. P. Dutton, 1968. Pp. vii, 140. $4.50.

Fiedler, Fred E. *A Theory of Leadership Effectiveness.* New York: McGraw-Hill, 1967. Pp. vii, 310. $8.50.

Freeman, Lucy. *The Cry for Love.* New York: Macmillan, 1969. Pp. x, 245. $5.95.

Gallup, Howard F. *An Invitation to Modern Psychology.* New York: The Free Press, 1969. Pp. xiii, 252. $2.95.

Greenwald, Anthony G., Brock, Timothy C., and Ostrom, Thomas M. *Psychological Foundations of Attitudes.* New York: Academic Press, 1968. Pp. xii, 407. $12.50.

Gurowitz, Edward M. *The Molecular Basis of Memory.* Englewood Cliffs, N. J.: Prentice-Hall, 1969. Pp. 92. $2.75.

Gustaitis, Rasa. *Turning On.* New York: Macmillan, 1969. Pp. xviii, 326, $6.95.

Hare, A. Paul, and Blumberg, Herbert H., eds. *Nonviolent Direct Action: American Cases, A Social Psychological Analysis.* Washington, D. C.: Corpus Publications, 1969. Pp. viii, 575. $10.00.

Hart, Jane, and Jones, Beverly. *Where's Hannah? A Handbook for Parents and Teachers of Children with Learning Disorders.* New York: Hart, 1968. Pp. 256. $8.50.

Hayter, Alethea. *Opium and the Romantic Imagination.* Berkeley: University of California Press, 1968. Pp. 388. $7.50.

Heath, Helen. *Elementary Statistics: Organized and Simplified.* Springfield, Ill.: Charles C Thomas, 1968. Pp. xiii, 146. $8.75.

Hellmuth, Jerome, ed. *Exceptional Infant.* Vol. 1, *The Normal Infant.* Seattle: Special Child Publications, 1967. Pp. 568. $12.50.

Hellmuth, Jerome, ed. *Learning Disorders.* Vol. 1–3. Seattle: Special Child Publications, 1965, 1966, 1969. Pp. 406, 422, 605. $4.85, $4.85, $12.50.

Hicken, Victor. *The American Fighting Man.* New York: Macmillan, 1969. Pp. 259. $6.75.

Hoffman, Martin. *The Gay World.* New York: Basic Books, 1968. Pp. ix, 212. $5.95.

Honigmann, John J. *Personality in Culture.* New York: Harper & Row, 1967 Pp. xii, 495.

Indik, Bernard P., Berrien, F. Kenneth, eds. *People, Groups, and Organizations.* New York: Teachers College Press, 1968. Pp. xii, 275. $6.00.

Insko, Chester A. *Theories of Attitude Change.* New York: Appleton-Century-Crofts, 1967. Pp. viii, 374. $7.50.

Isaacson, Robert L. *The Neuropsychology of Development: A Symposium.* New York: Wiley, 1968. Pp. 177. $8.95.

Jacobson, Edmund. *Biology of Emotions.* Springfield, Ill.: Charles C Thomas, 1967. Pp. xvii, 211. $7.50.

Jervis, George A., ed. *Expanding Concepts in Mental Retardation.* Springfield, Ill.: Charles C Thomas, 1968. Pp. v, 262.

Johnson, George L. *Certain Psychological Characteristics of Retarded Readers and Reading Achievers.* New York: Vantage Press, 1968. Pp. xiii, 137. $5.00.

Joncich, Geraldine. *The Sane Positivist: A Biography of Edmund L. Thorndike.* Middletown, Conn.: Wesleyan University Press, 1968. Pp. 634. $12.50.

Jones, G. Seaborn. *Treatment or Torture.* New York: Barnes & Noble, 1968. Pp. x, 565. $15.00.

Jordan, Nehemiah. *Themes in Speculative Psychology.* New York: Barnes & Noble, 1969. Pp. 259. $6.75.

Jung, John. *Verbal Learning.* New York: Holt, Rinehart and Winston, 1968. Pp. ix, 212. $5.95.

Kimble, Gregory A. *Foundations of Conditioning and Learning.* New York: Appleton-Century-Crofts, 1967. Pp. xii, 696. $8.75.

Kirscht, John P., and Dillehay, Ronald C. *Dimensions of Authoritarianism.* Lexington: University of Kentucky Press, 1967. Pp. xvi, 168.

Klein, Donald C. *Community Dynamics and Mental Health.* New York: Wiley, 1968. Pp. xvi, 224. $7.95.

Köhler, Wolfgana. *Werte und Tatsachen.* Berlin: Springer Verlag, 1968. Pp. 299. Paperback, $14.50.

Konorski, Jerzy. *Conditioned Reflexes and Neuron Organization.* (Reprinted from 1948.) New York: Hafner, 1968. Pp. 277. $11.00.

Kraepelin, Emil. *Lectures on Clinical Psychiatry.* (Facsimile of 1904 ed.) New York: Hafner, 1968. Pp. 308. $12.50.

Kurweil, Z. E. *Anxiety and Education.* New York: Thomas Yoseloff, 1968. Pp. 201. $6.00.

Lanyon, Richard I. *A Handbook of MMPI Group Profiles.* Minneapolis: University of Minnesota Press. 1968. Pp. v, 78. $7.50.

Lawson, Chester A. *Brain Mechanisms and Human Learning.* Boston: Houghton Mifflin, 1967. Pp. xiii, 126. $4.50.

Lazarus, Richard S. *Patterns of Adjustment and Human Effectiveness.* New York: McGraw-Hill, 1969. Pp. xiii, 280. $9.50.

Lepp, Ignace. *Death and Its Mysteries.* New York: Macmillan, 1968. Pp. xxv, 194. $4.00.

Lidz, Theodore. *The Person.* New York: Basic Books, 1968. Pp. 608. $10.00.

Lindgren, Henry Clay, ed. *Contemporary Research in Social Psychology.* New York: Wiley, 1969. Pp. ix, 653. Cloth, $9.95; paperback, $5.95.

Lindgren, Henry Clay. *An Introduction to Social Psychology.* New York: Wiley, 1969. Pp. 378. $8.95.

Lindgren, Henry Clay. *The Psychology of College Success.* New York: Wiley, 1969. Pp. 141. Cloth, $4.95; paperback, $2.45.

Litwin, George H., and Stringer, Robert A., Jr. *Motivation and Organizational Climate.* Boston: Harvard Business School, 1968. Pp. xiii, 214. $6.00.

Lloyd-Jones, Esther M., and Rosenau, Norah, eds. *Social and Cultural Foundations of Guidance.* New York: Holt, Rinehart and Winston, 1968. Pp. xi, 643. $6.95.

Lomax, Alan. *Folk Song Style and Culture.* Washington, D. C.: American Association for the Advancement of Science, 1968. Pp. xix, 363. $16.75.

Lorang, Sister Mary Corde (Maryknoll). *Burning Ice: The Moral and Emotional Effects of Reading.* New York: Charles Scribner's Sons, 1968. Pp. xii, 303. $7.95.

Lunzer, E. A., and Morris, J. F. *Development in Human Learning.* New York: American Elsevier, 1969. Pp. xii, 487.

McCartney, William. *Olfaction and Odours.* Berlin: Springer Verlag, 1968. Pp. 249. $9.00.

Madison, Bernice Q. *Social Welfare in the Soviet Union.* Stanford, Calif.: Stanford University Press, 1968. Pp. xxvi, 298. $8.50.

Martin, Kevin. *How to Go to High School or College by Mail.* New York: Frederick Fell, 1969. Pp. xii, 254. $5.95.

Maynert, Theodore. *Psychiatry.* New York: Hafner, 1969. Pp. 285. $13.50.

Meerloo, Joost. *Suicide and Mass Suicide.* New York: E. P. Dutton, 1968. Pp. vi, 153. $4.95.

Milton, Theodore. *Approaches to Personality.* New York: Pitman, 1968. Pp. xii, 539. Cloth, $10.75; paperback, $5.75.

Miner, John B. *Personnel and Industrial Relations.* New York: Macmillan, 1969. Pp. 562. $9.95.

Mishler, Elliot G., and Waxler, Nancy E. *Interaction in Families.* New York: Wiley, 1968. Pp. x, 436. $11.95.

Myrdal, Alva. *Nation and Family.* Cambridge: The M.I.T. Press, 1968. Pp. xxiii, 441. $3.45.

Neff, Walter S. *Work and Human Behavior.* New York: Atherton Press, 1968. Pp. 280. $8.50.

Neimark, E. D., and Estes, W. K. *Stimulus Sampling Theory.* San Francisco: Holden-Day, 1967. Pp. v, 681. $11.50.

Norman, D. A. *Memory and Attention.* New York: Wiley, 1969. Pp. v, 201. $8.95.

Oetting, Eugene, R., and Thornton, George C., III. *Exercises in Psychological Testing.* New York: Harper & Row, 1968. Pp. 229. $3.95.

Parker, Buelah. *Mental Health In-Service Training.* New York: International Universities Press, 1968. Pp. 132. $4.00.

Parry, John. *The Psychology of Human Communication.* New York: American Elsevier, 1968. Pp. ix, 248. $6.00.

Parsons, P. A. *The Genetic Analysis of Behavior.* New York: Barnes & Noble, 1967. Pp. ix, 174, $4.50.

Peeling, Kitty A. *The Unwanted.* New York: Vantage Press, 1969. Pp. 67. $2.75.

Peter, Laurence F., and Hull, Raymond. *The Peter Principle.* New York: William Morrow, 1969. Pp. ix, 248. $6.00.

Polatin, Phillip, and Philtine, Ellen C. *How to Develop a Well-Developed Personality.* New York: Essandess Special Editions, 1968. Pp. 150. $1.50.

Pranger, Robert J. *Action, Symbolism, and Order.* Nashville, Tenn.: Vanderbilt University Press, 1968. Pp. ix, 225. $6.95.

Pye, Lucian W. *The Spirit of Chinese Politics.* Cambridge: The M.I.T. Press, 1968. Pp. xxii, 255. $2.95.

Ravenette, A. T. *Dimensions of Reading Difficulties.* New York: Pergamon Press, 1968. Pp. xiv, 102. Cloth, $3.25; paperback, $2.00.

Ray, Isaac. *Mental Hygiene.* (Facsimile of 1863 ed.) New York: Hafner, 1968. Pp. 338. $10.50.

Resnik, H. L. P., ed. *Suicidal Behaviors: Diagnosis and Management.* Boston: Little, Brown and Company, 1968. Pp. xxvii, 536. $15.00.

Rhenman, Eric. *Industrial Democracy and Industrial Management.* The Netherlands: Royal Van Gorcum, 1968. Pp. 174.

Rhine, J. B., and Brier, Robert. *Parapsychology Today.* New York: Citadel Press, 1968. Pp. 268. $6.00.

Roberts, Leigh M., Greenfield, Norman S., and Miller, M. H., eds. *Comprehensive Mental Health.* Madison: University of Wisconsin Press, 1968. Pp. v, 339. $10.00.

Roby, Thornton B. *Small Group Performance.* Chicago: Rand McNally, 1968. Pp. 357. $7.00.

Rothney, John W. M. *Methods of Studying the Individual Child.* Waltham, Mass.: Blaisdell, 1968. Pp. 144. $1.95.

Schroder, Harold M., Driver, M., and Streufert, S. *Human Information Processing.* New York: Holt, Rinehart and Winston, 1967. Pp. vii, 224. $7.95.

Scrimshaw, Nevin S., and Gordon, John E., eds. *Malnutrition, Learning, and Behavior.* Cambridge: The M.I.T. Press, 1968. Pp. xiii, 566. $12.50.

Silverman, Charlotte. *The Epidemiology of Depression.* Baltimore: The Johns Hopkins Press, 1968. Pp. xviii, 184. $7.50.

Simon, Herbert J. *The Sciences of the Artificial.* Cambridge: The M.I.T. Press, 1969. Pp. x, 123. $5.95.

Smith, Frank, and Miller, George A. *The Genesis of Language.* Cambridge: The M.I.T. Press, 1966. Pp. ix, 400. $2.95.

Snitzer, Herb. *Living at Summerhill.* New York: Collier Books, 1968. Pp. 142. $1.95.

Speier, Hans. *Social Order and the Risks of War.* Cambridge: The M.I.T. Press, 1969. Pp. viii, 497. $3.95.

Spence, Kenneth W., and Spence, Janet Taylor. *The Psychology of Learning and Motivation.* Vol. 2. New York: Academic Press, 1968. Pp. 249. $10.50.

Spicer, E. C., ed. *British Medical Bulletin: Computing in Medicine.* London: Medical Department, the British Council, 1968. Pp. 276. Paperback, $6.50.

Spuhler, J. N. *Genetic Diversity and Human Behavior.* Chicago: Aldine, 1967. Pp. x, 291. $7.50.

Stewart, Walter A. *Psychoanalysis: The First Ten Years, 1888 to 1898.* New York: Macmillan, 1967. Pp. 225. $5.95.

Stocking, George W., Jr. *Race, Culture, and Evolution.* New York: The Free Press, 1968. Pp. xvii, 380. $10.00.

Sunder, S. D. *Hypnosis: A Clinical Study.* New York: Asia Publishing House, 1968. Pp. 155. $4.25.

Szasz, Thomas S. *Law, Liberty, and Psychiatry.* New York: Collier, 1963. Pp. xi, 281. $2.45.

Tenenbaum, Samuel. *A Psychologist Looks at Marriage.* Cranbury, N. J.: A. S. Barnes, 1968. Pp. 405. $6.95.

Tenon, Mathew. *Belief of the Stranger.* New York: A. S. Barnes, 1961, 1968. Pp. 117. $4.00.

Turner, Merle B. *Philosophy and the Science of Behavior.* New York: Appleton-Century-Crofts, 1967. Pp. xviii, 539.

Ullman, Leonard P., and Krasner, Leonard. *A Psychological Approach to Abnormal Behavior.* Englewood Cliffs, N. J.: Prentice-Hall, 1969. Pp. vii, 687.

Vandenberg, Steven G. *Progress in Human Behavior Genetics.* Baltimore: The Johns Hopkins Press, 1968. Pp. xi, 356. $12.50.

Van Over, Raymond, and Oteri, Laura. *William McDougall: Explorer of the Mind.* New York: Helix Press, 1967. Pp. 319. $8.50.

Vinacke, W. Edgar. *Foundations of Psychology.* New York: American Book, 1968. Pp. xvi, 847.

Vinacke, W. Edgar. *Readings in General Psychology.* New York: American Book, 1968. Pp. x, 560.

Warr, Peter B., and Knapper, Christopher. *The Perception of People and Events.* New York: Wiley, 1968. Pp. viii, 445. $10.00.

Weber, Jean-Paul. *The Psychology of Art.* New York: Delacorte Press, 1969. Pp. 187. Cloth, $6.00; paperback, $1.95.

Welford, A. J. *Fundamentals of Skill.* New York: Barnes & Noble, 1968. Pp. 426. $11.25.

Willems, Edwin P., and Raush, H. L., eds. *Naturalistic Viewpoints in Psychological Research.* New York: Holt, Rinehart and Winston, 1969. Pp. 294. $7.50.

Wilshire, Bruce. *William James and Phenomenology.* Bloomington: Indiana University Press, 1968. Pp. 251. $10.50.

Wing, Herbert. *Tests of Musical Ability and Appreciation.* New York: Cambridge University Press, 1968. Pp. 98. $4.75.

Wisdom, John. *Other Minds.* Berkeley: University of California Press, 1968. Pp. vii, 265. $2.25.

Wright, George N., Smits, Stanley J., Butler, Alfred J., and Thoreson, Richard W. *A Survey of Counselor Perceptions.* Madison: University of Wisconsin Press, 1968. Pp. viii, 65. Paperback.

Yeomans, K. A. *Statistics for the Social Scientist.* Vol. I, *Introducing Statistics.* Vol. II, *Applied Statistics.* Baltimore: Penguin Books, 1968. Pp. vii, 258; vii, 397.

Zytowski, Donald G. *Vocational Behavior Readings in Theory and Research.* New York: Holt, Rinehart and Winston, 1968. Pp. x, 486. $7.95.

THE AMERICAN
JOURNAL OF PSYCHOLOGY

Founded in 1887 by G. STANLEY HALL

VOL. 82 SEPTEMBER 1969 NO. 3

A MODIFIED REPLICATION OF COOK'S INVARIANCE-HYPOTHESIS EXPERIMENT

By E. E. RUMP, University of Adelaide, Australia

A standard object is positioned in the frontoparallel plane at a distance D from the observer; its frontal size is S. If σ is the perceived objective size and δ the perceived distance, the invariance hypothesis (IH) predicts:

$$\sigma/\delta = S/D. \qquad [1]$$

Several difficulties arise in testing the IH. In particular, measures of σ and δ must necessarily be indirect. Because of this, validity may be doubtful, *e.g.* when physical measures for a comparison object are used as σ and δ. Moreover, indirect measures involving kinesthetic[1] or proprioceptive[2] judgments are liable to have high error variance. Errors of measurement may also lead to bias, even if it were true that "variations due to inaccuracies of measurement alone would be randomly distributed."[3] Certainly, constant errors of position, order, and direction of adjustment may occur in estimates of σ and δ.[4] Partly because of this problem, at least one scaling constant may be introduced into the IH equation, which

* Received for publication August 30, 1968.

[1] W. C. Gogel and H. W. Mertens, Perceived size and distance of familiar objects, *Percept. mot. Skills*, 25, 1967, 213–225.

[2] W. C. Gogel, B. O. Hartman, and G. S. Harker, The retinal size of a familiar object as a determiner of apparent distance, *Psychol. Monogr.*, 71, 1957 (No. 13, Whole No. 442).

[3] W. Epstein, J. Park, and A. Casey, The current status of the size-distance hypothesis, *Psychol. Bull.*, 58, 1961, 511.

[4] The error of the standard, as described by J. F. Wohlwill, The development of 'overconstancy' in space perception, in L. P. Lipsitt and C. C. Spiker (eds.), *Advances in Child Development and Behaviour*, 1, 1963, 265–312. Also see L. Karlin, The time error in the comparison of visual size, this JOURNAL, 66, 1953, 564–573. Note further that, for example, an effect of direction of movement on distance judgments was reported by T. E. Weckowicz, and R. Hall, Distance constancy in schizophrenic and non-schizophrenic mental patients, *J. clin. Psychol.*, 16, 1960, 274.

may then be regarded as giving a correlational prediction:[5]

$$\sigma/\delta = k \cdot S/D \, . \qquad [2]$$

However, the doubtful validity associated with measures of σ and δ remains. Cook has suggested a method which overcomes this problem.[6] He has derived equations from simultaneous equations of the above kind, such that the perceptual terms and the constant k have been eliminated algebraically. In describing the experimental method and equation, a simpler notation than Cook's is used here to aid understanding. A measure of physical size is represented by S and a measure of physical distance from O by D. Subscripts are used to denote the three objects viewed: (st) represents the standard, (s) the size variable, and (d) the distance variable. Thus, $S_{(st)}$ represents the physical size of the standard. Other constants for the particular experimental conditions are $S_{(d)}$, $D_{(st)}$, and $D_{(s)}$. Size matches and distance matches are also represented in this way. In Cook's method, O makes three types of matching judgment in turn. The size variable is adjusted so that it looks equal in objective size to the standard, to give the value $S_{(s)}$. The distance variable is adjusted so that it looks as far away as the standard, giving $D_{(d)}$. Finally, the size variable is adjusted again, but in this case so that it looks equal in objective size to the distance variable; this second size match is denoted by $S'_{(s)}$. For this last judgment, the distance variable must be set at distance $D_{(d)}$ for Cook's prediction equation to be strictly applicable. The three judgments are repeated over a series of n trials, and means are obtained to give $\bar{S}_{(s)}$, $\bar{D}_{(d)}$ and $\bar{S}'_{(s)}$ for each O.

Equation 2 may be applied to the situations when these matching judgments have been completed, and by algebraic combination and substitution, this equation may be derived:

$$\bar{S}'_{(s)} = \{S_{(d)} \cdot D_{(st)} \cdot \bar{S}_{(s)}\} / \{S_{(st)} \cdot \bar{D}_{(d)}\} \, . \qquad [3]$$

This is Cook's Equation 9 in the simplified notation. The derivation is given by Cook, and it is convincing (apart from obvious printing errors). However, it is important that the nature of the equation should be recognized. First, it is not an alternative to the IH as represented by Equation 2 but, rather, is a strict application of the

[5] E. E. Rump, The relationship between perceived size and perceived distance, *Brit. J. Psychol.*, 52, 1961, 111.

[6] M. L. Cook, An exact test of the size-distance invariance hypothesis, this JOURNAL, 79, 1966, 568–575. I am grateful to this author for discussing his derivations with me.

IH to the experimental conditions outlined. Second, the perceptual terms σ and δ have been eliminated, and all terms in Equation 3 represent *physical* sizes or distances, both conceptually and in their operational identification.

In order to demonstrate the relationship between Equation 3 and the more conventional application of the IH, the former may be rearranged by algebraic transposition of terms:

$$\bar{S}_{(s)} = S_{(st)} \cdot \left\{ \frac{\bar{D}_{(d)}}{D_{(st)}} \right\} \cdot \left\{ \frac{\bar{S}'_{(s)}}{S_{(d)}} \right\}. \tag{4}$$

Equation 4 predicts, then, that the size match set equal to the standard is the product of the standard's actual size, its distance match relative to actual distance, and the size match set equal to the distance variable relative to the latter's actual size. Now if $\bar{S}'_{(s)} = S_{(d)}$, which would be the case if the sizes of both size and distance variables are seen veridically, Equation 4 reduces to the special case:

$$\bar{S}_{(s)} = S_{(st)} \cdot \left\{ \frac{\bar{D}_{(d)}}{D_{(st)}} \right\}. \tag{5}$$

This special case predicts that the size match to the standard is proportional to its distance match and has often been used in testing the IH. It is therefore termed the conventional equation.

The foregoing demonstration has interesting theoretical consequences. Conventionally, Equation 5 has been derived from the basic Equation 2 by substituting $k = 1$, $\sigma = \bar{S}_{(s)}$, and $\delta = \bar{D}_{(d)}$. The last two substitutions imply an assumption that for size and distance variables viewed under optimal conditions, perceptual and physical scales are identical. On the other hand, the above derivation of Equation 5 as a special case of Cook's equation does not depend on such a sweeping assumption. It depends only on two very limited assumptions (which may be checked empirically): that the mean size match to the distance variable at distance $D_{(d)}$ is without error, and that adjustments to the size variable do not affect its perceived distance. Cook's formulation may be seen, then, as placing the conventional Equation 5 on a less speculative basis. Moreover, Equation 4 is its extension to cases in which it can not be assumed that the size match to the distance variable is without error.

There is a further development which arises from the need to make estimates of population parameters from the sampled matching judgments. Cook has shown that an unbiased estimate of the product of expectations of $D_{(d)}$ and $S'_{(s)}$ is obtained for Equation

4 if a correction term is introduced for any correlation between the sampled values of $D_{(d)}$ and $S'_{(s)}$ over the n trials. The precise version of the prediction equation, for use when the sample of trials is small, is then:[7]

$$\bar{S}_{(s)} = S_{(st)} \cdot \left\{ \frac{\bar{D}_{(d)} \cdot \bar{S}'_{(s)} - \frac{1}{n} \text{ covar } D_{(d)} \cdot S'_{(s)}}{D_{(st)} \cdot S_{(d)}} \right\}. \qquad [6]$$

For convenience, the right-hand side of Equation 6 is referred to as Y; the basic prediction for the experiment described below, then, was that the variable $\bar{S}_{(s)}$ is equal to or at least highly correlated with its predicted value, Y.[8] It may be noted that the covariance correction term in Equation 6 has a very small effect in practice. In the experiment described below, its value as a proportion of $\bar{D}_{(d)} \cdot \bar{S}'_{(s)}$ was only .01%, a negligible effect.

Having described the nature of Cook's method and its relation to the conventional method of testing the IH, certain criticisms must be noted. It has been pointed out that two additional assumptions are required: that the standard remains unchanged in perceived size while the size matches and distance matches are obtained, and that the standard is seen as upright.[9] However, no O in the experiment reported here remarked on a change in the size of the standard during any trial, nor were sloping stimuli reported, so that these possible sources of error are of little interest. A more relevant criticism concerns Cook's method of reporting his results. He reported that 10 of 11 Os gave a negative error of prediction, but he did not discuss the size of the discrepancy. The deviation from exact prediction in Cook's experiment could, then, have been quite small and due to some bias in the measures involved. It can not be concluded that the IH is completely invalidated by his results. It is more appropriate to ask the *degree* of predictive value afforded by Equation 6 and whether it is an improvement on the conventional Equation 5. In particular, there could be a high correlation between $\bar{S}_{(s)}$ and Y, even if exact equality of the two terms is not obtained for any O.

[7] This equation is derived from Cook's Equation 16 by setting Cook's parameter "D" equal to the predicted value of zero and by algebraic transposition.

[8] A convenient computational formula for Y is:

$$Y = S_{(st)}[S_{(d)} \cdot D_{(st)} \cdot n]^{-1}[(n+1) \cdot \bar{D}_{(d)} \cdot \bar{S}'_{(s)} - n^{-1} \cdot \Sigma(D_{(d)} \cdot S'_{(s)})].$$

[9] R. Britten, The size-distance invariance hypothesis: A reply to Cook, this Journal, 80, 1967, 640–642.

In 1961, Rump concluded that there was little evidence for large correlations across *Os* between size judgments and distance judgments in line with the IH.[10] However, since that time a number of such positive correlations have been reported.[11] It is possible that individual differences in the perceived size of the distance variable might occur, in which case the use of Equation 6 might give a higher correlation than the conventional Equation 5. An experiment was therefore designed to compare the accuracy of prediction from the two equations, using three different standard objects.

METHOD

Three groups of *Os* were tested, one group for each standard. The standards were a matchbox front cover (as in Cook's experiment), 5.34 cm. in height and 3.60 cm. in width; an oversized cigarette-box front cover (from a Viscount 'king size' display pack), 17.00 cm. in height and 11.70 cm. in width; and as a control, a white rectangular card of the same size as the cigarette box. The height of the 'cigarette box' was 204% that of a normal one.

All *Os* and *Es* were men, to minimize nonexperimental communications.[12] All but two *Os* were tested by an assistant,[13] and all procedures were followed from a typescript. The first *O*'s session was treated as a training session, with both *Es* present; the results were not analyzed. To avoid the slight changes in the position of the standards that would have occurred if they were randomly ordered, 12 *Os* were tested with the card, then 12 others with the cigarette box, and then 12 more with the matchbox. Four additional *Os* were tested with the card, two by the author and two by the assistant as a check on order and experimenter effects.

Cook's method was followed, with some modifications. Cook's standard and distance variable appeared to lie in a darkened chamber. In the present experiment the objects were illuminated by two fluorescent ceiling lights, to avoid any extreme judgments which might have occurred if *Os* suspected 'experimental trickery.' Moreover, the apparatus was placed in an ordinary office (unfamiliar to *Os*), with shuttered windows along the right side allowing some slight daylight to enter.

Apparatus. All stimulus objects were perpendicular to *O*'s line of sight, had their longer sides vertical, and were seen against a matte-black background. The distance variable, placed centrally, was a normal playing card, with $S_{(d)} =$ 8.84 cm. It was attached to a vertical black rod, itself fixed into a sliding saddle on a 4-m. calibrated metal optical bench. This extended horizontally from near *O*'s viewing position to the far end of the room. The size variable

[10] Rump, *op. cit.*, 116.

[11] *E.g.* .77 and .74 (W. W. Blessing, A. A. Landauer, and M. Coltheart, The effect of false perspective cues on distance- and size-judgments: An examination of the invariance hypothesis, this JOURNAL, 80, 1967, 250–256); also .47 and .74 (J. C. Baird, Retinal and assumed size cues as determinants of size and distance perception, *J. exp. Psychol.*, 66, 1963, 155–162).

[12] R. Rosenthal, Covert communication in the psychological experiment, *Psychol. Bull.*, 67, 1967, 356–367.

[13] N. F. Riedl, whose work is gratefully acknowledged.

was a 2.8-cm.-wide white rectangle, positioned 35° to the right of d and at a distance of 100 cm. It could be raised through a slot in a large screen to vary its height. The size variable was in full view of O so as to make feasible the assumption that adjustment of its size would not affect its perceived distance.

Each standard was screened and positioned 20° to the left of d, with $D_{(st)} =$ 275 cm. Viewing was through a reduction tube of 2.53-cm. internal diameter and 35.5-cm. length. The reason for using reduced-cue conditions is as follows. If all stimuli are seen under optimal conditions, judgments are probably nearly veridical over short distances, in which case Equation 2 and its derivations, including Equation 6, necessarily hold true. In order to properly test for predictive value, some attempt must be made to induce erroneous judgment; only then does successful prediction indicate a strong verification of the IH.

The distance variable was also screened, and viewed through a similar reduction tube, the two tubes being joined to give a common viewing aperture. Again, the reduction tube was used to increase error of judgment, because only when $S'_{(s)} \neq S_{(d)}$ is any distinction possible between the predictions of the conventional Equation 5 and Equation 6. The centers of the aperture and of all stimulus objects were 120 cm. from floor level. Position of O's head was controlled using a chin support and forehead rest, so that small turns of the head in either direction allowed O to observe each object with his right eye.

The reflected light from the direction of the stimulus objects was measured at the eye's position for each session, using a Lunasix meter and calibration charts. The mean values converted to apostilb units were: for the size variable set at 17 cm., 16.5; for the distance variable at 275-cm. distance, 0.8; for the matchbox, 0.4; for the cigarette box, 0.5; and for the white card, 1.5.

Observers. The *Os*, 40 undergraduates naïve with respect to the IH, had a mean age of 19 years, 10 months. Three other students were not used as *Os* because pretests indicated poor acuity.

Procedure. The O was led to his chair behind a screen with his eyes closed. Earmuffs were provided to reduce auditory cues, and a shield over the left eye to ensure that judgments were monocular. For each trial, values of $S_{(s)}$, $D_{(d)}$, and $S'_{(s)}$ were determined in that order. Order and procedures were standardized to avoid 'artifactual' individual differences. Each value was the average of an increasing and a decreasing adjustment. The size of s was increased by O's pulling a cord and decreased by his pulling another cord, while maintaining the viewing position. The distance of d was similarly adjusted by O. Both pairs of cords were directly connected to the variables by means of fixed pulleys, whose effects were demonstrated immediately before trials began. This method of adjustment allowed quick but precise judgments to be repeated for nine trials, with little intervention by E.

The $S_{(s)}$ values were read by E from a scale on the back of the screen; the $D_{(d)}$ values were read from the optical-bench scale, with a constant added for the distance between the viewing position and the nearer end of the scale. During the $S_{(s)}$ judgments, the end of the d reduction tube was covered; and during the $S'_{(s)}$ judgments, the end of the st tube was covered. All distances in the experiment were recorded to the nearest 1 mm., and all sizes to the nearest 0.2 mm.

The starting positions of the variables for increasing and decreasing judgments were symmetrical about the zero-error value, so as to minimize bias

due to 'central tendency.' The positions were: for d, 112 and 438 cm.; and for s matched with the card or cigarette box, 1 and 33 cm., with the matchbox, 1 and 9.7 cm., and with the playing card, 1 and 16.7 cm. 'Objective' instructions were used for $S_{(s)}$ and $S'_{(s)}$, emphasizing that the actual heights should be equal if the objects compared were placed face to face. For the distance match, Os were instructed to adjust the playing card until "it looks exactly the same distance from your eye as the matchbox [or cigarette box or white card]."

To obtain an estimate of subjective uncertainty of judgments, at the end of the session each O was asked to indicate "the limits within which you are 90% sure that the size of the *st* lies." These upper and lower limits were indicated by adjustment of the size variable. The difference between the two settings gave the 'range of uncertainty' for $\bar{S}_{(s)}$. The range of uncertainty corresponding to $\bar{D}_{(d)}$ judgments was also obtained in a similar way by adjustment of the distance variable.

Results

Initial treatment. The first trial for each O was treated as practice and was not analyzed. The distance of d set by E during the determination of $S'_{(s)}$ for each of the 320 analyzed trials was checked. In two trials, an error in setting d of just over 1 cm. was made; these trials were rejected, and all calculations were based on seven trials for these two Os. The main statistics were checked by use of a CDC 6400 computer after initial manual calculation. The values over trials of $\bar{S}_{(s)}$, $\bar{D}_{(d)}$, $\bar{S}'_{(s)}$, $\Sigma S'_{(s)} \cdot D_{(d)}$, and Y were found for each O.

The Y against $\bar{S}_{(s)}$ plot for the white card was inspected, and the four final Os were found to show results similar to the original Os; these 16 Os were therefore combined for further analysis. Group-mean values of the main statistics are given in Table I. It is emphasized that these statistics, including Y, were first calculated for each O separately before averaging over the groups. In all cases in which statistics are given in the text for the matchbox, cigarette box, and white card, they are quoted in that order. Levels of significance are one-tailed for predicted directional effects.

Prediction of group means. Equation 6 may be used to predict the size match to the standard, $\bar{S}_{(s)}$, from the other matches, $\bar{D}_{(d)}$ and $\bar{S}'_{(s)}$. The predicted value is the right-hand side of the equation, termed Y. The mean error of prediction ($\bar{S}_{(s)} - Y$) should equal zero. The obtained values for this error of prediction were +.01 cm., −1.28 cm., and −0.49 cm. for the three standards. None of these values was significantly different from the predicted zero using two-tailed tests: the t values were .04 1.92, and 0.34, with 11, 11, and 15 df respectively.

There was thus no consistent bias of the kind found by Cook,

TABLE I

GROUP MEANS AND CORRELATIONS FOR THREE STANDARDS

Measure	Notation	Matchbox	Cigarette box	White card
Number of *O*s	N	12	12	16
Actual size of standard	$S_{(st)}$	5.34	17.00	17.00
Actual distance	$D_{(st)}$	275.0	275.0	275.0
Group means				
Size match with standard	$\bar{S}_{(s)}$	4.99	11.56	16.68
Distance match with standard	$\bar{D}_{(d)}$	324.9	183.1	272.3
Size match with distance variable	$\bar{S}'_{(s)}$	6.93	9.99	9.10
Correction term	$\frac{1}{n}$ covar $D_{(d)} \cdot S'_{(s)}$	−.30	−.19	.01
Predicted size match using Equation 6	Y	4.99	12.84	17.17
Predicted size match using Equation 5	$S_{(st)} \cdot \bar{D}_{(d)} / D_{(st)}$	6.31	11.32	16.83
Linear correlations				
Between $\bar{S}_{(s)}$ and Y		.94	.83	.74
Between $\bar{S}_{(s)}$ and $\bar{D}_{(d)}$		.40	.60	.61

Note: All measures in cm.

and indeed, the error in prediction was negligible with the matchbox similar to the one he used. In Cook's experiment the mean error of prediction was about +0.62 cm., so that although consistent in direction (for 10 of 11 *O*s), the error was quite small.[14] It may have been larger than for the matchbox in the present experiment because of some slight judgmental bias, such as that discussed in the following section. Or the different conditions of illumination in the present experiment may have facilitated more accurate use of the relative retinal-size cue.

The errors of prediction found with Equation 6 may also be compared with those using the conventional Equation 5, *i.e.* with values of $(\bar{S}_{(s)} - S_{(st)} \cdot \bar{D}_{(d)} / D_{(st)})$. However, in comparing accuracy of prediction for the two equations, it is more appropriate to consider the magnitude of the error of prediction for each *O* regardless of sign. The average values for the modulus error of prediction, for the three standards respectively, were as follows: using Equation 6, 0.66 cm., 1.85 cm., and 4.42 cm.; and using the conventional Equation 5, 1.66 cm., 1.95 cm., and 5.52 cm. The two equations did not differ reliably in their predictive accuracy for the cigarette box and the card, but Equation 6 gave significantly less error of prediction than

[14] The value of mean $(\bar{S}_{(s)} - Y)$ for Cook's experiment was calculated from his published data and from the value of $D_{(st)}$ given in a personal communication as 130.3 in.

Equation 5 for the matchbox standard using a two-tailed, related-samples t test ($p < .05$).

This result was to be expected from an examination of the size matches to the distance variable, $\bar{S}'_{(s)}$ (see Table I). The value for the matchbox group was 6.93, which is markedly less than for the other groups and less than the actual size of d, 8.84 cm. Equation 6 takes into account any such underestimation of the size of the distance variable.[15] It therefore gave a better prediction in this case than Equation 5, which assumes $\bar{S}'_{(s)} = S_{(d)}$, as shown earlier. These results illustrate the point that Cook's derivation does allow fairly accurate predictions, but they also demonstrate that if there is no large bias in judgment of the size of the distance variable, the conventional equation gives a quite accurate prediction also.

The group means may be examined further with regard to the familiar-size hypothesis. This predicts that mean $\bar{S}_{(s)}$ for the oversized cigarette box should be smaller than for the control card. It follows from Equation 6 that mean Y should also be correspondingly smaller for the cigarette box. The mean $\bar{S}_{(s)}$ value for the cigarette box was 11.56 cm., which is significantly less than 16.68 cm. for the control card ($p < .05$) and significantly less than its actual size of 17 cm. ($p < .01$). The mean Y value was 12.84 cm., which is significantly less than 17.17 cm. for the control card ($p < .05$).[16] The prediction variable Y was thus reliably reduced for the oversized familiar object, to a value approximately corresponding to its mean size match. In contrast, for the control card both mean $\bar{S}_{(s)}$ and mean Y were close to the actual size $S_{(s,c)}$, which was well within the 95% confidence limits of both variables. Neither mean $\bar{S}_{(s)}$ nor mean Y values were reduced as far as the normal size (8.34 cm.) of a cigarette box of the type used, so that in this sense the judgments show a compromise between mnemonic and visual cues.

Correlations across observers. The accuracy of prediction from Equations 6 and 5 may also be assessed by the linear correlations between $\bar{S}_{(s)}$ and its predicted value. These correlations are given in Table I and are graphed in Fig. 1–3. The correlations between $\bar{S}_{(s)}$ and Y predicted by Equation 6 were high, and significant for each

[15] The underestimation, and the effect of the $S'_{(s)}$ variable, would probably have been less if the playing card had been viewed under less reduced conditions (H. R. Schiffman, Size-estimation of familiar objects under informative and reduced conditions of viewing, this JOURNAL, 80, 1967, 229–235).

[16] Because of unequal variances, a correction to df was used in assessing significance in these cases (W. L. Hays, *Statistics for Psychologists*, 1963, 322).

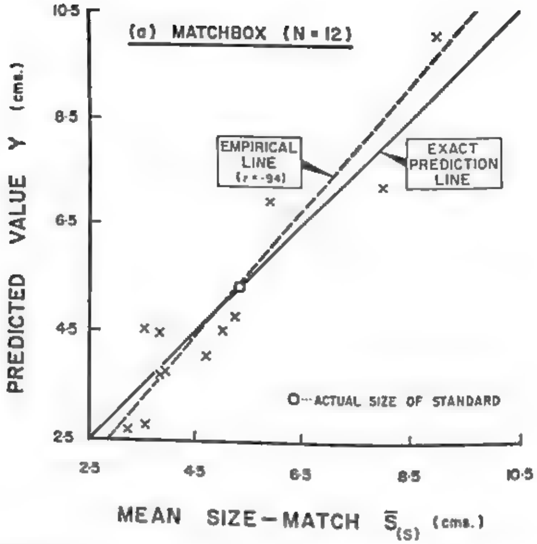


FIG. 1. RELATIONSHIP BETWEEN SIZE MATCH AND ITS PREDICTED VALUE: MATCHBOX STANDARD

group ($p < .01$). These correlations were higher than those between $\bar{S}_{(s)}$ and $\bar{D}_{(d)}$ predicted by the conventional Equation 5 (other terms in this equation being constants). Again, the greater accuracy of Equation 6 was pronounced only for the matchbox.

An additional indication of the reliability of the correlations between $\bar{S}_{(s)}$ and Y was provided by pilot studies which gave the following inter-*O* correlations for four independent groups: with an oversized matchbox as standard, .95 ($N = 8$); with the oversized cigarette box, .61 ($N = 9$); and with white cards, .64 ($N = 9$) and .90 ($N = 8$). Under the experimental conditions described, the IH allows good, though not exact, prediction of relationships between size matches and distance matches, across *O*s, using Cook's method and Equation 6.

The contribution of the size match to the distance variable (the $\bar{S}'_{(s)}$ component of Y) to the correlation between Y and $\bar{S}_{(s)}$ is of interest. That the contribution was important is shown by the correlations between $\bar{S}_{(s)}$ and $\bar{S}'_{(s)}$: .89 ($p < .01$), .69 ($p < .01$), and .37 for the three standards.[17] These correlations illustrate the relevance of taking the individual's idiosyncrasy in making size matches into account when predicting $\bar{S}_{(s)}$, by use of the $\bar{S}'_{(s)}$ variable as

[17] These correlations between size judgments for different objects are similar to those found using a drawing method (A. C. McKennell, Visual size and familiar size: Individual differences, *Brit. J. Psychol.*, 51, 1960, 30).

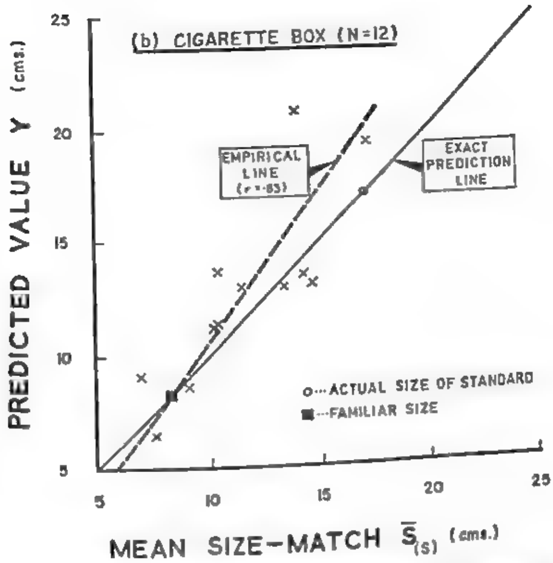


FIG. 2. RELATIONSHIP BETWEEN SIZE MATCH AND ITS PREDICTED VALUE: CIGARETTE-BOX STANDARD

Equations 4 and 6 require. It might be suggested that the slight variations in illumination for different *O*s spuriously increased the correlations between $\bar{S}_{(s)}$ and Y. This possibility is contradicted by the correlations with illumination partialed out, using the mean photometric reading for each *O*. These partial correlations between $\bar{S}_{(s)}$

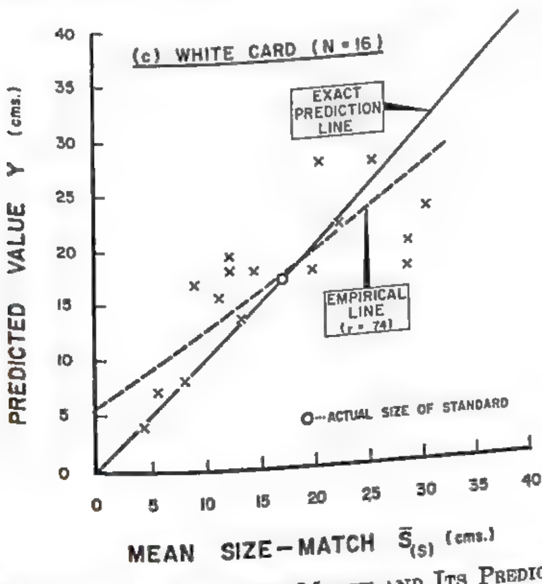


FIG. 3. RELATIONSHIP BETWEEN SIZE MATCH AND ITS PREDICTED VALUE: WHITE-CARD STANDARD

and Y were .92, .81, and .74, so that variations in illumination had a negligible effect.

Possible reasons for the deviation of individual *O*s' judgments from the exact prediction line ($\bar{S}_{(s)} = Y$) may be inferred. Less accurate prediction might be expected for *O*s who are very uncertain of their judgments. To obtain an overall index of uncertainty, each *O*'s 'range of uncertainty' scores for size and for distance were taken as proportions of his $\bar{S}_{(s)}$ and $\bar{D}_{(d)}$ values respectively. These proportions were standardized for each group separately, and the two standard scores added to give an overall uncertainty score for each *O*. The correlation between these scores and the magnitude of the error of prediction ($\bar{S}_{(s)} \sim Y$) was .72 ($p < .01$) for the cigarette-box group. The corresponding correlations for the other groups were small and not significant; the relationship may therefore occur only when there are conflicting cues, such as occur with a standard which is familiar but of unusual size.

In connection with the large errors of prediction for certain *O*s, inspection of Fig. 3 shows that there were in the card group three *O*s in particular who had large $\bar{S}_{(s)}$ values without commensurately high Y values. Their low Y values were due to their distance matches being less than required by the IH. A possible explanation of these anomalous judgments is that the distance settings for these three *O*s were near the upper limit of the distance scale, and sampled $D_{(d)}$ values were therefore skewed towards the center of the scale. This negative skew reduced $\bar{D}_{(d)}$, and thus the obtained Y values for these *O*s, to give the ($\bar{S}_{(s)}$, Y) points below the prediction line. The other groups did not have such high distance-match values and therefore did not show this 'end of scale' effect. Some bias of measurement like this might possibly account for the consistent deviation from prediction found by Cook.

The predicted slope. From Equation 6, the function Y on $\bar{S}_{(s)}$ should not only be linear but also have a slope of unity, apart from sampling error. To estimate the slope, a line of best fit is required. A regression solution of the first kind was not appropriate, because there was error on both variables. Moreover, the usual maximum-likelihood or least-squares solutions were not possible, because estimates of the two error variances and their intercorrelation independent of the main regression effect can not be obtained.[18] Instead, the regression

[18] F. S. Acton, *Analysis of Straight-Line Data*, 1959, 135. The two error variances for $\bar{S}_{(s)}$ and Y can not be assumed equal to allow the usual least-squares solution, since Y includes *two* terms with error variance, $\bar{S}'_{(s)}$ and $\bar{D}_{(d)}$.

coefficient for the main axis of the "tolerance ellipse" was used.[19] Resulting estimates of slope for the three standards were 1.16, 1.38, and 0.69. These slopes, together with the mean $\bar{S}_{(s)}$, mean Y points, define lines to fit each plot as shown in Fig. 1. While the slopes differ somewhat from unity, it is noteworthy that the average of the three sampled values, weighted by the *df* on which each is based, was approximately unity as predicted, viz. 1.03.

Correlations across trials. Variation of judgment from trial to trial is usually treated as entirely due to error if conditions are constant. However, it is possible that there might be some nonrandom component to intertrial variance and, in particular, that some degree of IH relationship might exist. To investigate this possibility, Y values were calculated for every trial, though the covariance correction term was neglected since it is assumed constant for all trials of any individual *O*. Because each trial consisted of only two adjustments for each of the three variables and there were only eight trials used (seven for two *O*s), the sampled correlations between $S_{(s)}$ and Y across trials were small and variable. Therefore, the intertrial correlation between $S_{(s)}$ and Y was calculated for each *O* and then averaged over all 40 *O*s, after normalizing correlations by Fisher's Z transformation.[20] The mean Z so obtained was .147, which corresponds to a value of .15 for the estimated population intertrial correlation. This was small, but positive as predicted and significantly different from zero with a critical ratio of 2.07 ($p < .05$).

In order to assess the adequacy of the approximation afforded by the conventional Equation 5, the mean intertrial correlation between $S_{(s)}$ and $D_{(d)}$ was also calculated in a similar manner. The mean correlation was .05, which is not significantly different from zero. Equation 6 was thus superior to Equation 5 for the prediction of intertrial relationships.

The relativity of error. It is quite generally found in psychophysical studies of magnitude estimation that error and uncertainty are relative to the magnitude of the estimation. This has been termed Ekman's law.[21] Several aspects of this study appear to confirm the

[19] Konrad Diem (ed.), *Documenta Geigy: Scientific Tables*, 6th ed., 1962, 181.

[20] There was a degree of heterogeneity among the intertrial correlations, but since it was not significant ($\chi^2 = 52$, with 39 *df*), use of the overall mean Z was permissible (Diem, *op. cit.*, 62).

[21] S. S. Stevens, A metric for the social consensus, *Science*, 151, 1966, 530–541.

'law' to some degree. The error of prediction ($\bar{S}_{(s)} \sim Y$) was relative to the magnitude of Y, with linear correlations across Os of .63 ($p < .05$), .68 ($p < .01$), and .44 ($p < .05$) for the three standards. The 'range of uncertainty' for size matches correlated across Os with the magnitude of $\bar{S}_{(s)}$: .13, .80 ($p < .01$), and .39 for the three standards. Similarly, the 'range of uncertainty' for distance matches correlated with the magnitude of $\bar{D}_{(d)}$: .53 ($p < .05$), .93 ($p < .01$), and .38.

Uncertainty of judgment is also indicated by intertrial variance. Because of skew in the distribution of variances, Spearman's rank correlation was used. The correlations between $\bar{S}_{(s)}$ and its intertrial variance were .51 ($p < .05$), .64 ($p < .05$), and .70 ($p < .01$) for the three standards. Similarly, the rank correlations between Y and its intertrial variance (neglecting the covariance correction term) were .78 ($p < .01$), .83 ($p < .01$), and .81 ($p < .01$). These various measures of error, uncertainty, and variability of judgment therefore correlated positively, and in most cases quite highly, with the magnitude of judgment.

Differences between groups in variability. It might be expected that variability of judgment would depend on the standard used. Variability should be least for a familiar object of standardized size giving veridical mnemonic and visual cues (the matchbox), greater for an object giving nonveridical mnemonic cues (the cigarette box), and greatest for an object giving no visual cues from surface pattern nor any mnemonic cues (the white card).

This order was indeed found with respect to the three measures of variability examined, for both the size-match and predicted size-match variables (see Table II). For $\bar{S}_{(s)}$, the card had a significantly greater inter-O standard deviation than the other two objects ($p < .01$ in each case, using F ratios). For Y, the matchbox had a significantly smaller inter-O standard deviation than the cigarette box and card ($p < .05$ and .01 respectively). The matchbox also tended to have smaller intra-O standard deviations than the other two objects, both for $S_{(s)}$ and for Y ($p < .01$ in all four cases, using Mann-Whitney tests).[22] The cigarette box tended to have smaller intra-O standard deviations than the card for the Y variable ($p < .05$). The *max.F* values indicating the spread of intra-O variances showed the same trend, though no significance test for differences between *max.F* values is known by the author.

[22] Mann-Whitney tests were used rather than pooling the intra-O variance owing to the heterogeneity of the standard deviations concerned.

TABLE II

MEASURES OF VARIABILITY FOR THREE STANDARDS

Measure	Variable	Matchbox	Cigarette box	White card
Inter-*O* standard deviation†	$\bar{S}_{(s)}$	1.77	2.97	8.39
	Y	2.04	3.88	6.38
Median intra-*O* standard deviation	$\bar{S}_{(s)}$	.27	.53	.67
	Y	.33	.68	1.09
Spread of intra-*O* variances (*max.F*)	$\bar{S}_{(s)}$	7.8	11.6	161.1‡
	Y	32.8	98.1	164.2‡

Note: All measures in cm.
† Sample values are given, but population estimates were used in the calculation of *F* ratios to test differences between groups.
‡ Calculated for the first 12 *O*s so that *max.F* ratios are directly comparable for the three groups.

The consistency of these results suggests that a primary effect of an increase in cues, and perhaps also in the ability to utilize cues, may be to decrease variability of perceptual judgments. Unfortunately, this effect was somewhat confounded in the present study with the Ekman's-law effect, so that confirmation is required. There are, in fact, several findings which help to confirm the interpretation offered. Ittelson refers to "the greater coefficients of variation" in distance matches for unfamiliar as compared with familiar objects, though his standard deviations were not entirely consistent in this respect.[23] Standard deviations were lower with binocular vision compared with monocular vision with the same standard, for both size and distance matches.[24] Prior training in making distance judgments reduced both inter-*O* and intra-*O* variance in subsequent judgments.[25] On the other hand, groups with low intelligence, brain damage, or schizophrenia, *i.e.* groups presumably less able to utilize cues, had greater variability in size judgments compared with normals, although, admittedly, differences were not significant in all cases.[26] The consistency of this evidence supports the postulated effect of cues and their utilization on variability, despite the lack of investigations specifically designed to test that hypothesis.

[23] W. H. Ittelson, Size as a cue to distance: Static localization, this JOURNAL, 64, 1951, 54–67.

[24] W. Epstein, Nonrelational judgments of size and distance, this JOURNAL, 79, 1966, 121, Table I.

[25] E. J. Gibson, R. Bergman, and J. Purdy, The effect of prior training with a scale of distance on absolute and relative judgments of distance over ground, *J. exp. Psychol.*, 50, 1955, 97–105.

[26] V. Hamilton, Size constancy and intelligence: A re-examination, *Brit. J. Psychol.*, 57, 1966, 322, Table 2; M. Wyke, Alterations of size constancy associated with brain lesions in man, *J. Neurol. Neurosurg. Psychiat.*, 23, 1960, 256, Table III; R. Stannard, G. Singer, and R. Over, The effect of informational feedback on size judgments of schizophrenic patients, *Brit. J. Psychol.*, 57, 1966, 329–333.

Summary

A prediction equation was derived from Cook's equations, which were derived from the invariance hypothesis relating perceived size to perceived distance. The prediction equation relates size and distance matches for a standard stimulus and allows for relative error in size matches to the distance variable. To test the predicted relationship, groups of *Os* made judgments with respect to three different standards. The relationship was confirmed with respect to differences between *Os* and between standards; intertrial differences also gave a small predicted correlation. The derived equation gave more accurate predictions than the conventional III equation in some, but not all, cases. Evidence was also found for Ekman's law of the relativity of uncertainty in judgments and for the effect of reducing cues on the variability of judgments.

EFFECTS OF *D*-AMPHETAMINE ON ACQUISITION, PERSISTENCE, AND RECALL

By Paul M. Hurst, Robert Radlow, Nicholas C. Chubb, and Sallyann K. Bagley,
Institute For Research, State College, Pennsylvania

Although amphetamines have frequently been administered in learning situations, their effects upon learning itself are not conclusively established. In some situations, the drugs increase the rate at which performance improves during the learning phase. Eysenck, Casey, and Trouton observed that *d*-amphetamine improved performance during the learning of a pursuit-rotor task.[1] Franks and Trouton obtained a similar improvement during the learning of associations between stimulus lights and response buttons.[2] These experiments suggest that amphetamines increase learning, but they are not conclusive. As Weiss and Laties' review of these studies concluded, "simply promoting a greater rate of improvement in proficiency is insufficient proof, in view of the numerous effects of amphetamine on motor performance."[3]

Weitzner performed a more compelling demonstration.[4] He observed facilitation by amphetamine during the learning of paired associates having high association value and low intralist competition. (He also found no significant drug effect under high com-

* Received for publication April 12, 1968. The research reported was performed on Contract Nonr 4423(00) with the Physiological Psychology Branch of the Office of Naval Research. The authors wish to thank Dr. Albert L. Ingram, Jr., and his staff at the Pennsylvania State University Health Center for their valuable cooperation in *S*-screening operations. The authors also thank Dr. Alberto DiMascio for furnishing the high-competition word lists from his experiment. Drugs were donated by Smith Kline and French Laboratories. Dr. Radlow is now at San Diego State College, San Diego, California; Dr. Chubb is with the Ritenour Health Center, Pennsylvania State University, University Park.

[1] H. J. Eysenck, S. Casey, and D. S. Trouton, Drugs and personality: II, The effect of stimulant and depressant drugs on continuous work, *J. ment. Sci.*, 103, 1957, 645–649.

[2] C. M. Franks and D. Trouton, Effects of amobarbital sodium and dexamphetamine sulfate on the conditioning of the eyeblink response, *J. comp. physiol. Psychol.*, 51, 1958, 220–222.

[3] B. Weiss and V. G. Laties, Enhancement of human performance by caffeine and the amphetamines, *Pharmacol. Rev.*, 14(1), 1962, 1–36.

[4] M. Weitzner, Manifest anxiety, amphetamine, and performance, *J. Psychol.*, 60, 1965, 71–79.

petition, *i.e.* when many response words had high association values for each stimulus word, a fact which becomes important below.) Weitzner's experiment was cited by Laties and Weiss as "the first unambiguous demonstration that amphetamine can improve verbal learning. . . ."[5] If one defines learning as recall of associations, this is true. Yet even this demonstration does not fulfill the paradigm for a conclusive learning experiment. Weitzner himself concluded simply that "the drug significantly facilitates performance" in at least one type of paired-associate learning.

The paradigm for demonstration of learning is, rather, illustrated in Elliott's classical study in which manipulation of food incentive produced large differences in rats' maze-running performance, with the preferred-incentive group showing progressively superior performance during the acquisition phase.[6] When the food incentives were reversed for the two groups, the performance differential was also reversed, and much more abruptly than would be expected from the previous learning curves. The group that was shifted to nonpreferred incentive actually showed a sharp decline in performance. Thus, the appropriate paradigm for measuring learning achievement necessitates removal of the agent whose effects on learning are being assessed. With a drug, this is possible only in a delayed test, since a drugged subject can not immediately be rendered undrugged. Consequently, the present study measured not only *acquisition* but also *persistence* of that material during a session of delayed recall and relearning. Related questions concerned *recall facilitation, i.e.* whether a drug present during such a delayed recall test would affect recall and relearning of the material, and *drug interaction, i.e.* whether any such effect would be independent of the drug condition during the first (learning, or acquisition) session. There might, for example, be an interaction between the two drug treatments due to common cue properties during learning and recall.

The present study was also concerned with whether the effects of the drug are dependent upon the degree of intralist *competition* per se or upon the degree of item *difficulty*. By increasing 'arousal,' amphetamines may augment the interference effects of competitive associations. This interpretation is supported not only by Weitzner's

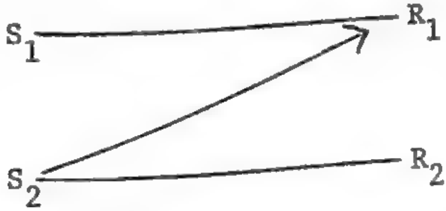
[5] V. G. Laties and B. Weiss, Performance enhancement by the amphetamines: A new appraisal, University of Rochester, Department of Radiation Biology and Biophysics, 1966.

[6] M. H. Elliott, The effect of change of reward on the maze performance of rats, *Univ. Calif. Pub. Psychol.*, 4, 1928, 19–30.

results (those mentioned parenthetically above) but also by DiMascio's finding that a depressant drug (meprobamate) in fact improved learning of paired associates having high intralist competition.[7] The Weitzner and DiMascio results do not, however, prove that high intralist competition itself is responsible for enhanced learning by depressant drugs and lack of enhancement by amphetamines. DiMascio did not include any low-competition lists for cross comparison; and although Weitzner did find a contrast between the two types of list, it is possible that the crucial variable there was not interference from high intralist competition, but simply task difficulty. Therefore, since Weiss and Laties found that more difficult or complex tasks generally tend to show less enhancement from amphetamines, the present study attempted to separate the role of difficulty from that of competition per se by manipulating difficulty along other, independent dimensions, *i.e.* usage frequency and association value. Finally, during the experiment a within-session serial effect appeared, and the mechanisms that might account for this were also considered.

Method

Task. Twelve lists of paired associates were used, each consisting of 14 pairs of words. Four lists were identical to the high-competition lists used by DiMascio—lists which were constructed according to the paradigm of Mednick and DeVito:[8]



Here, S_1 and R_1 have a high probability of free association, S_2 and R_2 have a very low probability, and S_2 has a higher probability of association with R_1 than has S_1. The remaining eight lists, all of low internal competition, included two lists each of these four types: (1) low association, low frequency (LALF); (2) low association, high frequency (LAHF); (3) high association, low frequency (HALF); and (4) high association, high frequency (HAHF). Thus, the experimental design was hierachical, with association value and usage frequency nested in the low-competition condition.

[7] A. DiMascio, Drug effects on competitive-paired associate learning: Relationship to and implications for the Taylor manifest anxiety scale, *J. Psychol.*, 56, 1963, 89–97.

[8] S. A. Mednick, A learning theory approach to schizophrenia, *Psychol. Bull.*, 55, 1958, 316–327.

For the eight low-competition lists, words of selected association values and usage frequencies were obtained from Palermo and Jenkins,[9] Gerow and Pollio,[10] Deese,[11] and Thorndike and Lorge.[12] The high-frequency (HF) stimulus and response words had median usage frequencies greater than 100 per million. The low-frequency (LF) stimulus words had a median usage frequency of 1–2 per million, and the low-frequency response words a median of 50–100 per million. The high-association (HA) pairs had a median free association of 0.155, and the low-association (LA) pairs zero free association per thousand. (It was also found that DiMascio's original lists had median association values of 0.13 for the connected high-association pairs, with zero for the connected low-association pairs and 0.39 for the nonconnected, 'competitive' associations. His stimulus words had a median usage frequency of 50–100 per million, and his response words a median usage frequency of over 100 per million.) All lists were recorded on magnetic tape.

Subjects and drug treatment. University students (men) who were at least 21 years old were recruited as paid volunteers. Those who gave informed consent to the experiment were then screened by the medical supervisor. Of those who completed the experiment, the raw median age was 21. The treatments employed were *d*-amphetamine sulfate (dex) and placebo. The dosage of the former was adjusted to the two-thirds power of bodyweight, with the constant of proportionality adjusted to assign a 14-mg. dose to a 70-kg. *S*. This dosage had been previously found to be well tolerated and generally more effective than smaller dosages.

Session I. Ss were randomly assigned to the four groups (treatment sequences) shown in Table I, with 17–18 *Ss* per group. The session took place on an evening; *Ss* had been instructed to eat normally before they came. The capsules were ingested at 6:30 P.M., and the latent period was occupied by writing on an assigned topic, by mood checklists, and by instructions. Before explaining the test procedure, *E* explained how the *Ss'* payments would be calculated. Each *S* was guaranteed a base payment of $2.00 for each session completed, with a bonus for completing both sessions; the bonus to be based on total score, with the top performer during the two test sessions receiving $42.00 and each of the other *Ss* receiving a fraction of that top bonus equal to the quotient of his score (total number correct) divided by the top score.

The task requirements were then described, and a practice test of 10 word pairs was administered. The main test began at 7:50 and lasted until 9:25, thus occupying the interval between 80 and 175 min. after capsule ingestion. The lists were presented as follows: A tape recording first presented the 14 word pairs of a given list. Then, in changed order, the stimulus words were given. Each was followed after 4 sec. by the sound of a horn, then after 1 more sec. by the appropriate response word. *S* was told to write the first three

[9] D. S. Palermo and J. J. Jenkins, *Word Association Norms: Grade School through College*, 1964.

[10] J. R. Gerow and H. R. Pollio, Word association, frequency of occurrence, and semantic differential norms for 360 stimulus words, University of Tennessee, Department of Psychology, Technical Report No. 1, April 1965.

[11] J. Deese, Free association norms for 599 words, Johns Hopkins University, n.d.

[12] E. L. Thorndike and I. Lorge, *The Teacher's Word Book of 30,000 Words*, 1944.

TABLE I

EXPERIMENTAL DESIGN

Session	Group 1 ($N = 17$)	Group 2 ($N = 17$)	Group 3 ($N = 17$)	Group 4 ($N = 18$)
I. (Trials 1–3) Acquisition	Placebo	Placebo	Dex	Dex
II. (Trials 4) Delayed recall and (Trials 5–6) Relearning	Placebo	Dex	Placebo	Dex

Defining contrasts

1. Acquisition:	Groups $(1 + 2)$ vs. $(3 + 4)$, Session I, Trials 1–3
2. Persistence	
Delayed recall:	Groups $(1 + 2)$ vs. $(3 + 4)$, Session II, Trial 4
Relearning	Groups $(1 + 2)$ vs. $(3 + 4)$, Session II, Trials 5–6
3. Recall facilitation	
Delayed recall:	Groups $(1 + 3)$ vs. $(2 + 4)$, Session II, Trial 4
Relearning:	Groups $(1 + 3)$ vs. $(2 + 4)$, Session II, Trials 5–6
4. Interaction	
Delayed recall:	Groups $(1 + 4)$ vs. $(2 + 3)$, Session II, Trial 4
Relearning:	Groups $(1 + 4)$ vs. $(2 + 3)$, Session II, Trials 5–6

Note: Persistence refers to availability of previously learned material for delayed recall, as a function of the treatment during original learning (Session I). Recall facilitation refers to the effect of the treatment present during the test of delayed recall and relearning (Session II), regardless of treatment during Session I.

letters of the response word before he heard the horn. At this signal, he was required to flip over the answer page for that word. This procedure, which was enforced by the use of monitors and a separate color-coded answer sheet for each word, ensured that the answer would be written before it was heard. After the initial reading of a list, it was presented three times for recall (Trials 1–3). After the last trial, a 30-sec. rest was given, during which a brief historical sketch was read, to minimize interlist interference. This was followed by the reading of the next list, and then by the three recall trials for that list, and so on through the 12 lists. The sequence of lists was counterbalanced with respect to the experimental variations in competition, usage frequency, and association value.

Session II. At the session that took place one week later, the same lists were again presented for recall in the same manner, but without the initial nontest reading of each list. Thus, the first trial of this session (Trial 4) with each list gave a measure of delayed recall, and the two subsequent presentations (Trials 5–6) measured relearning. The *Ss* had not been told that they would be retested on these same materials. The between-group contrasts used to assess component effects may be clarified by referring back to Table I. In evaluating the Session II results, it should be kept in mind that the present procedure is not the customary one. The results measure retention of material learned to a variable degree after a fixed amount of practice, rather than material learned to a fixed criterion after a variable amount of practice. The former criterion seemed preferable, since this study was based on the assumption that *d*-amphetamine may affect the relationship between learning and the expression of learning via performance. If this were true, any attempt to equate groups by a fixed criterion of initial 'learning' would be biased, since the learning criterion must be inferred from manifest performance.

Results

Acquisition. The first analysis concerned the acquisition phase (Session I), with results contrasted only on the basis of the drug treatment on that occasion (Defining Contrast 1, Table I). The result for all lists combined ('main effect' of drug) was not significant. However, the mixed-model analysis used to secure this result revealed interactions between drug treatment and list types that were significant, as discussed below. Hence, it was desirable to measure the drug influence as a separate, simple effect at each level of each list variable and at the combinations thereof. Accordingly, a series of further analyses was undertaken to reveal the drug effect separately in each of the 13 data subsets defined by the combinations of different levels of list variables, recalling that variations in association value and usage frequency were not feasible in the high-competition condition. This nested design yielded only a single, overall measure of drug effects under high competition, whereas the drug effects under low competition could be analyzed separately at different levels of association value and usage frequency. The results of these separate analyses are given in Table II.

Since only one of these 13 partially independent analyses showed even marginal significance ($p < 0.10$), one can not reject the null hypothesis for the effects of drug treatments. It is interesting to note that the most nearly significant effect obtained with a relatively small difference in treatment means (3.2%), whereas differences of 13% and 14% failed even to approach significance. This is because the larger differences were all obtained with various LA subsets, and the LA condition produced much greater variability than HA.

The list variables, of course, produced some highly significant effects: competition ($p < 0.001$), association value ($p < 0.001$), but not usage frequency ($p > 0.10$). Interactions between drug and list variables were assessed according to the mixed design they represented (Type I, Lindquist):[13] drug treatment was found to interact significantly with association value ($p < 0.01$), but not with competition ($p > 0.10$) or with usage frequency ($p > 0.10$). The interaction with association value could well be an artifact of a ceiling effect, since the HA lists were recalled almost perfectly by all *S*s and could hardly have shown the treatment difference found

[13] E. F. Lindquist, *Design and Analysis of Experiments in Psychology and Education*, 1953.

TABLE II

ACQUISITION: SESSION I: TRIALS 1-3

Lists	Correct responses (%)		Drug effect
	Dex	Placebo	$\frac{\text{Dex \%} - \text{Placebo \%}}{\text{Placebo \%}} \times 100\%$
High-competition	63.5	62.3	1.9
Low-competition			
LALF	50.1	44.3	13.1
LAHF	47.6	41.5	14.7
HALF	93.5	90.6	3.2*
HAHF	93.0	93.5	−0.5
Total LA	49.2	42.9	14.7
Total HA	93.2	92.0	1.3
Total LF	71.8	67.5	6.4
Total HF	70.6	67.5	4.6
Total low-competition	71.2	67.3	5.8
Total for first six lists	66.9	65.2	2.6
Total for last six lists	70.4	66.3	6.2
Total of all lists	68.6	65.8	4.3

Note: The list types were given in counterbalanced order, so the comparison of the first six lists is independent of the list-type variables.

* $p < 0.10$. This and other significance levels in the following tables refer to raw-score F tests. For main effects of drug treatment, $df = 1$ and 68. For all other main effects and interactions, $df = 1$ and 469.

in the LA lists. The directions of all of these effects and trends can be seen in the separate treatment means (Table II). Note that the Drug × Competition trend, while not significant, was in the direction predicted from Weitzner's results. This was overshadowed by the questionable but significant Drug × Association Value interaction (a drug effect of 14.7% with the combined LA lists, but only 1.3% with HA lists).

It should be noted that measurement precision was much higher for list variables and interactions than for the main effects of drug treatment, which is not surprising with a mixed design. The between-groups error mean square (for drug effects) was 224.6 raw-score units, whereas the within-groups error term (for list effects) was only 2.87.

Persistence. Delayed recall and relearning in Session II were first analyzed on the basis of drug treatment during Session I. This required combining groups who received different treatments during Session II itself, so that the effects of those treatments would be balanced and not contaminate the test of persistence (Defining Contrast 2, Table I). The dependent variables for measuring persistence were slightly different from those used to measure acquisition. With each of the various list types, separate analyses

were performed on results of Trial 4, the delayed-recall measure. Combined results from Trials 5 and 6 were used to measure relearning, since the amount relearned might reflect any savings effect due to persistence. Although the main effect of drug treatment was of borderline significance (see total of all lists, Table III), the presence of interactions again prompted supplementary separate analyses by list types and subtypes, the results of which also appear in Table III.

Note that the active drug produced its only significant effects under conditions of low competition, which implies the interaction between drug and list variables that is specified below. It was also determined that all three of the list variables produced highly significant ($p < 0.001$) main effects, although that of usage frequency was of small magnitude. The Drug × Association value interaction was significant ($p < 0.001$) only on Trials 5–6, probably because it was imprecisely measured on Trial 4 owing to the very low delayed-recall scores under low association. Again, this particular interaction could be an artifact of the ceiling effect for HA. The drug treatment did not interact significantly with any other list variable. There was a significant interaction ($p < 0.001$) not involving drug treatment, between association value and usage frequency. Once more, the ceiling effect may have been involved in the interaction. The direction of these list effects, interactions, and trends can be seen in Table III by comparing the drug's simple effects under various list conditions.

As with the Session-I data, the error term for drug effects was much larger than that for list-type variables. These error variances apply to both the present analyses and the following ones, since the latter differ only in using different defining contracts for treatment effects. For Trial 4 data, the between-groups error mean square was 11.6 raw-score units, whereas the within-groups error term (for list variables) was 2.22. For Trials 5 and 6 combined, the between-groups error term was 108.9 and the within-groups error term was 8.45.

Recall facilitation. To measure the effects of drug treatment during Session II, the two groups receiving *d*-amphetamine on that session were combined and contrasted with those receiving placebo (Defining Contrast 3, Table I). Since the two groups comprising either combination had each received a different treatment during acquisition, this contrast was independent of the differential persis-

TABLE III

PERSISTENCE: EFFECTS OF DRUG PRESENT DURING ACQUISITION ON SUBSEQUENT PERFORMANCE IN DELAYED RECALL AND RELEARNING: SESSION II

Lists	Delayed Recall: Trial 4			Relearning: Trials 5–6		
	Correct responses (%)		Drug effect	Correct responses (%)		Drug effect
	Dex	Placebo	$\frac{\text{Dex \%} - \text{Placebo \%}}{\text{Placebo \%}} \times 100\%$	Dex	Placebo	$\frac{\text{Dex \%} - \text{Placebo \%}}{\text{Placebo \%}} \times 100\%$
High-competition	23.2%	20.9%	11.0%	72.1%	70.0%	3.0%
Low-competition						
LALF	6.4	5.0	28.0	67.7	60.5	11.9
LAHF	3.9	2.9	34.5	59.6	51.2	17.0*
HALF	56.0	47.5	17.9**	97.3	96.8	0.5
HAHF	61.4	57.9	6.0	96.6	97.7	−1.1
Total LA	5.2	3.9	33.3	63.7	55.9	14.0*
Total HA	58.8	52.7	11.6**	97.0	97.1	−0.2
Total LF	31.2	26.2	19.1**	82.6	78.6	5.1
Total HF	32.7	30.4	7.6	78.1	74.5	4.8
Total low-competition	32.0	28.4	12.7**	80.3	76.5	5.0
Total for first six lists	30.8	29.0	9.7	75.8	74.2	2.2
Total for last six lists	27.5	22.7	21.1**	79.4	74.5	6.6*
Total of all lists	29.0	25.9	12.0*	77.6	74.3	4.4

Note: Dex and placebo here represent combined data from all *Ss* receiving those treatments in Session I.
* $p < 0.10$.
** $p < 0.05$.

tence effects of such prior treatments. Results are listed in Table IV. Once again, the drug effects were separately assessed at each level of every list variable or combination thereof.

Once more, the results suggested general superiority for the *S*s receiving *d*-amphetamine sulfate; as before, the negative differences were few and very small in size. However, the net margin was also rather modest and not significant. Whatever effect was exerted by the drug given just before Session II, it must have been slight and of lesser consequence than the delayed effect from the drug taken in Session I. The drug treatment applied in this session did not interact significantly with any of the list effects.

Interaction among drug treatments. Although interactions between drug treatment and list variables had already been observed, it remained to determine whether the drug variables interacted with each other, *i.e.* whether the effects of drug treatment during Session I and Session II were additive or not. For this purpose, the groups were partitioned in the remaining manner, so that the pair whose members had the same treatment on both occasions was contrasted with those having different treatments (Defining Contrast 4, Table I). This third orthogonal contrast yielded no differences of appreciable size or significance.

Discussion

Although *d*-amphetamine showed a generally facilitative trend at all stages of paired-associate learning, the strongest and the only significant effects were upon persistence of material learned under the active drug—regardless of whether or not the drug was also present when delayed recall was tested. This finding is the opposite of what would be expected if amphetamines had no true learning effect but acted instead to increase the performance potential at a given level of learning. Such an interpretation is also contradicted by the drug's failure to produce significant facilitation when present during Session II. Although the learning effect seems a real one, the results do not clearly distinguish the mechanisms involved. The moderating effect of competition, which can be related to theories of motivational arousal and interference, was only weakly supported. There were trends, both in acquisition and persistence, for *d*-amphetamine to produce more favorable effects with the low-competition paired associates. Also, there was an interaction between drug effect and association value, but this could well have

TABLE IV

RECALL FACILITATION: EFFECTS OF DRUG PRESENT DURING DELAYED RECALL AND RELEARNING: SESSION II

Lists	Delayed Recall: Trial 4			Relearning: Trials 5–6		
	Correct responses (%)		Drug effect	Correct responses (%)		Drug effect
	Dex	Placebo	$\frac{\text{Dex \%} - \text{Placebo \%}}{\text{Placebo \%}} \times 100\%$	Dex	Placebo	$\frac{\text{Dex \%} - \text{Placebo \%}}{\text{Placebo \%}} \times 100\%$
High-competition	23.2%	21.2%	9.4%	71.5%	70.4%	1.6%
Low-competition						
LALF	6.1	5.7	7.0	64.6	63.6	1.6
LAHF	3.6	3.2	12.5	56.6	54.3	4.2
HALF	51.4	52.5	−2.1	97.0	97.1	−0.1
HAHF	60.4	58.9	2.5	97.0	97.1	−0.1
Total LA	4.6	4.5	2.2	60.6	58.9	2.9
Total HA	55.9	55.7	0.4	97.0	97.1	−0.1
Total LF	28.8	28.9	−0.3	80.8	80.4	0.5
Total HF	32.0	31.1	2.9	76.9	75.7	1.6
Total low-competition	30.4	30.0	1.3	78.8	78.1	0.9
Total for first six lists	30.6	29.3	4.4	75.2	74.9	0.4
Total for last six lists	25.4	24.9	2.0	77.8	76.2	2.1
Total of all lists	27.9	27.1	3.0	76.4	75.5	1.2

Note: Dex and placebo here represent combined data from all *Ss* receiving those treatments in Session II.

been a ceiling effect due to the ease of the HA condition. There was no evidence of drug interaction with usage frequency. Consequently, one can not conclude that the observed drug effects support either the competition hypothesis or the difficulty hypothesis.

Another moderating influence observed was the within-test serial effect, such that the enhancement of later recall by *d*-amphetamine during acquisition was greater in the second half of the lists. This could be attributed to at least three different mechanisms. The first possibility is that the active drug's effect was still building up during the first half (80–127 min. after ingestion) and did not reach its peak until the second half (127–175 min. after ingestion). However, previous memory experiments of a different nature in this laboratory have yielded slightly better performance results with *d*-amphetamine in the earlier of two tests, although the latencies of this test were not the same as in the present study. (One was shorter and two were longer: see Hurst, Radlow and Bagley.[14]) This comparison with previous results also weakens the second possibility: that an amphetamine's enhancement of memory is mediated by reduction of boredom or fatigue, so that it is apparent only in material learned in the latter part of a session. The third possibility is that, in the present study, the drug somehow counteracted proactive inhibition involving intrusions from lists previously learned, as opposed to retroactive inhibition. Since not much is known about amphetamine's effects on learning in general, it is not surprising that little is known about specific effects upon proactive as opposed to retroactive inhibition. For the present, therefore, this third possibility must be considered a speculation to be further tested.

Summary

The experiment was designed to measure four component effects of *d*-amphetamine in a paired-associate learning task. The drug was compared with placebo to reveal (1) effects upon acquisition of new material, (2) persistence of material so learned, (3) effects of drug present during test of delayed recall, and (4) interactions between drug present during learning and drug present during test of delayed recall. Paired associates of varying degrees of usage frequency, association value, and intralist competition were employed. Sixty-nine college men were tested in two sessions, for

[14] P. M. Hurst, R. Radlow, and S. K. Bagley, Drug effects upon data processing as functions of storage and retrieval parameters, *Ergonomics*, in press.

learning and delayed recall respectively. Presence of *d*-amphetamine during learning resulted in significantly greater delayed recall of the low-competition lists and nonsignificantly greater delayed recall of high-competition lists. The same dosage given for the recall session had no significant effect upon delayed recall or relearning.

SUMMATION AND INTERACTION OF SUCCESSIVE MASKING STIMULI IN VISUAL PERCEPTION

By DONALD L. SCHURMAN and CHARLES W. ERIKSEN,
University of Illinois, Urbana

Most cases of backward and forward masking in visual perception appear explicable in terms of the lack of fine temporal resolution in the visual system. That is, successive stimuli presented with too short an interval between them are perceived as essentially simultaneous. The best-known example of this lack of resolution is Bloch's law, which states that within some critical duration, usually on the order of 100 msec., time and intensity of stimulation can be reciprocally interchanged without changing the visual effect. Such integration of energy over time by the visual system is apparently independent of the distribution of energy within this critical duration. Davy, for example, employed discrete pulses of light and found performance to be a function only of the total energy presented within the critical duration.[1]

The explanation of masking phenomena as due to lack of temporal resolution in the visual system implies that maximum effects should be obtained when the test and the masking stimuli are presented simultaneously. Most published results have been compatible with this requirement. Masking effects decrease as the interstimulus interval (ISI) increases. The few exceptions to this monotonic decreasing function[2] appear to be attributable to confounded effects such as apparent movement[3] or to effects of a delayed indicator designating the portion of the display for report.[4] The tend-

* Received for publication February 14, 1969. The research reported was supported by Public Health Service Research Grant MH-1206 and a Public Health Service Research Career Award, K6-MH-22,014.

[1] E. Davy, The intensity-time relation for multiple flashes of light in the peripheral retina, *J. opt. Soc. Amer.*, 42, 1952, 937–941.

[2] M. Alpern, Metacontrast: Historical introduction, *Amer. J. Optom.*, 29, 1952, 631–646; N. Weisstein, Backward masking and models of perceptual processing, *J. exp. Psychol.*, 72, 1966, 232–240.

[3] D. Kahneman, Time-intensity reciprocity in acuity as a function of luminance and figure-ground contrast, *Vision Res.*, 6, 1966, 207–215.

[4] C. W. Eriksen, J. F. Collins, and T. S. Greenspon, An analysis of certain factors responsible for nonmonotonic backward masking functions, *J. exp. Psychol.*, 75, 1967, 500–507.

ency of the visual system to integrate stimulation over durations of 100 msec. or so removes the mystery of how events act backward in time in perception. Instead, investigative efforts can be focused on understanding the effects and interactions of simultaneously presented stimuli. In the cases where the mask is a light flash, impairment in the perception of the target can be attributed to summation of the target luminance with the masking luminance and the attendant contrast reduction of the target form.[5] When the mask is a pattern or a ring, the impairment seems to be attributable to a change in the minimum acuity requirements or possible interactions of contours.[6]

Recently, however, Robinson has reported some rather anomalous results based on an experiment which used two masking stimuli instead of the usual single mask. The masks and the target were concentric disks of light, the first and smallest disk being the target, and the two larger disks the masking stimuli. When the largest disk (second mask) was presented 10 to 50 msec. after the first mask, detection of the target was improved over that when the target and the first mask were presented alone. Robinson interpreted these data as a demonstration of a disinhibition phenomenon, *i.e.* assumed the first masking flash to inhibit the perception of the target but the second mask, in turn, to inhibit the first mask and thus disinhibit the target.[7]

This disinhibition effect found by Robinson occurred when all three circular flashes were presented within a duration for which Bloch's law has been found to hold with light pulses.[8] According to Bloch's law the *S* should perceive a large disk of light which is brightest in the center and becomes progressively less bright toward the periphery. Since Robinson's stimuli were of different diameters, the possibility of simultaneous contrast, Mach bands, and apparent movement severely complicate the interpretation of his data and

[5] C. W. Eriksen, Temporal luminance summation effects in backward and forward masking, *Percept. & Psychophys.*, 1, 1966, 87–92; J. H. Thompson, What happens to the stimulus in backward masking? *J. exp. Psychol.*, 71, 1966, 580–586.

[6] D. G. Purcell, A. L. Steward, and W. N. Dember, Spatial effectiveness of the mask: Lateral inhibition in visual backward masking, *Percept. & Psychophys.*, 4, 1968, 344–346; C. W. Eriksen and J. F. Collins, Reinterpretation of one form of backward and forward masking in visual perception, *J. exp.* Psychol., 2, 1965, 351–352.

[7] D. N. Robinson, Visual disinhibition with binocular and interocular presentations, *J. opt. Soc. Amer.*, 58, 1968, 254–257.

[8] Davy, *loc. cit.*; M. Lichtenstein and R. Boucher, Minimum detectable dark interval between trains of perceptually fused flashes, *J. opt. Soc. Amer.*, 50, 1960, 461–466.

any predictions therefrom. Unfortunately, Robinson's experiment did not provide a control condition in which all three stimuli were simultaneously presented. Such a control, besides providing a possibility of disentangling some of the various effects that could be involved, would have also indicated whether asynchrony of the three flashes was essential to the effect.

Experiment I

An attempt was made to replicate Robinson's findings using similar stimuli but employing a forced-choice indicator in the hope of obtaining a more sensitive measure. Robinson had his *S*s use the numbers 1, 2, and 3 to report which of the disks of light they perceived on any given trial. In our procedure *S*s made a temporal forced choice to two sequences of stimulation as to which of the two sequences contained the small target flash. The target flash was always present in one of the two sequences and was in the first sequence 50% of the time in random order.

Subjects. Two graduate students, one of them a man, served as paid volunteers. Both *S*s had normal vision.

Apparatus. Six channels of a 10-channel projection tachistoscope were employed. This tachistoscope has been described in detail elsewhere, but in brief it consisted of a circular arrangement of 10 8-mm. projection systems with zoom lenses mounted so as to project to a common area on the rear of a ground-glass screen.[9] The individual lamps (Type F4T5/CWX) in each projector were timed by an Iconix timebase generator Model 6255, and two Iconix Model 6010 presets. The *S* viewed monocularly the front surface of the ground-glass screen through a visual tunnel 42 in. long. As viewed by *S* the equipment presented a circular field 6° of visual angle in diameter. The field was illuminated at approximately 0.02 mL. with a 0.5-mL. fixation cross of 15 min. of angle in the center.

The stimuli were constructed by carefully drilling holes in sheet metal. The sizes of the holes were such as to project a circle of light 14 min. in diameter for the target stimulus and 28 and 56 min. of angle for the first (F1) and the second masking stimulus (F2) respectively. These are the same sizes used by Robinson. From *S*'s viewpoint the target stimulus was 0.6 mL. in intensity while the two masking flashes were 1.2 mL.

Procedure. Prior to the experimental sessions each *S* was run for two practice sessions during which an exposure duration and luminance for the target disk was determined which yielded approximately 75% forced-choice detection accuracy when the target was presented simultaneously with F1. This determination was found to be necessary because *S*s yielded 100% correct detections when the target was presented for the same duration and at the same intensity as the mask, as Robinson did. Final determinations were at 20-msec.

[9] C. W. Eriksen, D. L. Schurman, and O. Richter, N-channel tachistoscopes, *Behav. Res. Meth. & Instrum.*, 1, 1969, 119–122.

duration for one S and 10-msec. for the other. The masks were of 20-msec. duration.

The four experimental conditions are schematized in Fig. 1. Condition A was a standard backward masking paradigm employing only the target disk and the F1 masking disk, the latter presented either simultaneously or at ISIs of 25 or 50 msec. from offset of target to the onset of the mask. Condition B was identical to Condition A except that F2, the larger masking disk, was presented simultaneously with F1. In Condition C, F2 always followed F1 by an ISI of

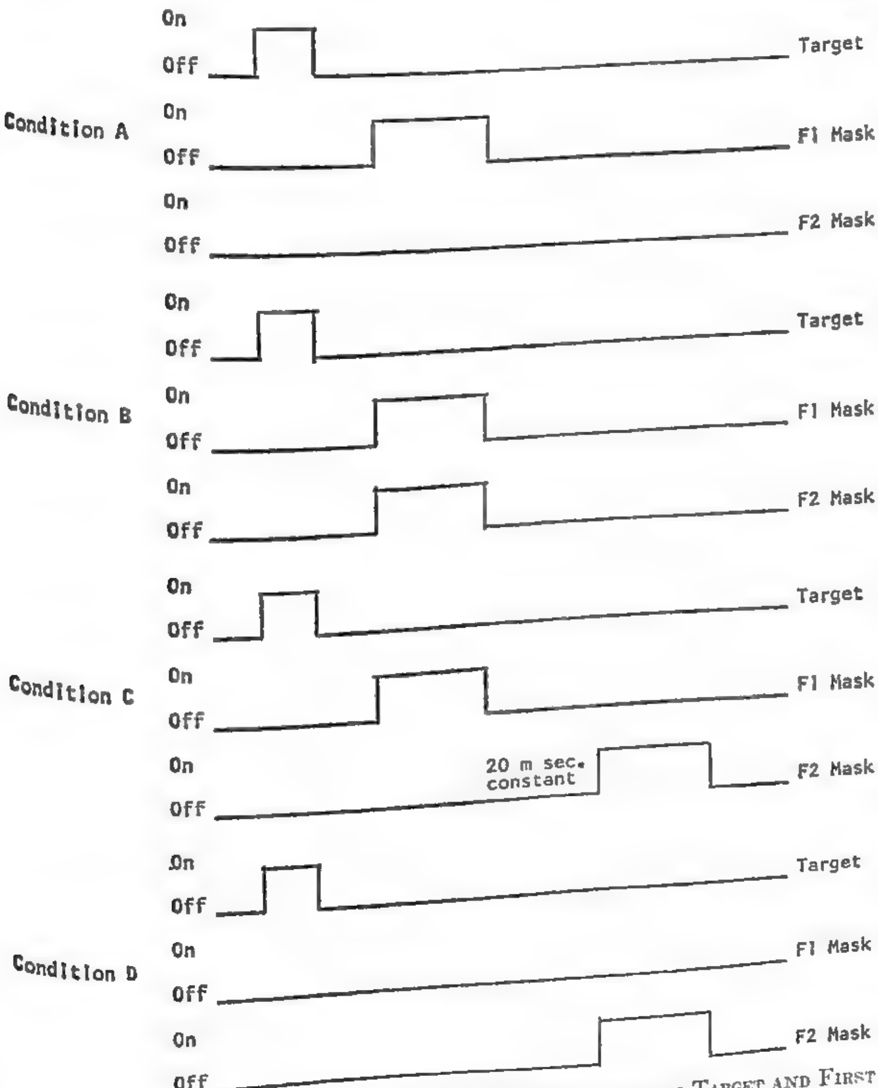


FIG. 1. SCHEMATIC DIAGRAM OF TEMPORAL RELATIONS BETWEEN TARGET AND FIRST AND SECOND MASKING FLASHES (F1 AND F2 MASKS): EXPERIMENT I

20 msec., and F1 again followed the target by 25 or 50 msec. ISI or was simultaneously presented. For Condition D, only F2 was employed and followed the target by varying ISIs as in Condition A. On a signal from *E*, *S* focused on the fixation point and, when it was clear and sharp, triggered the stimulation sequence. A trial consisted of two sequences spaced 1 sec. apart. After observing the two sequences *S* made a forced choice as to whether the target disk had been in the first or second sequence. The order in which the target was presented in the first or second sequence was random, with the restriction that it occur in each sequence equally often within a session. During an experimental session *S* made 24 judgments under each condition at each of the three ISI values. *Ss* served a total of four experimental sessions.

Results. The results for the two *Ss* were averaged and are shown in Fig. 2. Here the percentages of correct forced-choice responses have been plotted as a function of ISI between the target termination and onset of masking stimuli. For Conditions A and C, typical monotonic masking functions were obtained, with maximum masking occurring when target and masking stimuli were simultaneously presented. In Condition D, where only the target and the larger masking flash (F2) were presented, no masking was found. In this condition no observations were taken when these two stimuli occurred simultaneously. Instead, observations were collected when F2 followed the target at ISIs of 20, 65, and 90 msec. as it did in Condition C. The point of this arrangement was to ascertain the effect of the delayed F2 on the target when F1 did not intervene. As can be seen in Fig. 2, no masking occurred at any of these values.

Of major interest is the performance obtained under Conditions B and C relative to Conditions A and D. Under Condition B, F1 and F2 always occurred simultaneously at the designated values and the inhibition should have been maximal. Under Condition C, F1 followed the target at designated ISI values while F2 followed F1 at a constant 20-msec. ISI. The latter condition is the one under which Robinson reported the occurrence of disinhibition effects. Since performance under Conditions B and C at simultaneous ISI was at the chance level, it cannot be determined whether Conditions B and C differed in masking. At ISI values of 25 and 50 msec., however, both Conditions B and C yielded essentially the same masking as Condition A. There was a tendency for Condition B to give slightly less masking at the 25-msec. and 50-msec. ISIs, but the difference did not approach significance for either *S* ($p < .20$).

Discussion. The results of Experiment I did not provide any clear confirmation of Robinson's disinhibition effect. Condition C, which was highly similar to the condition under which Robinson obtained

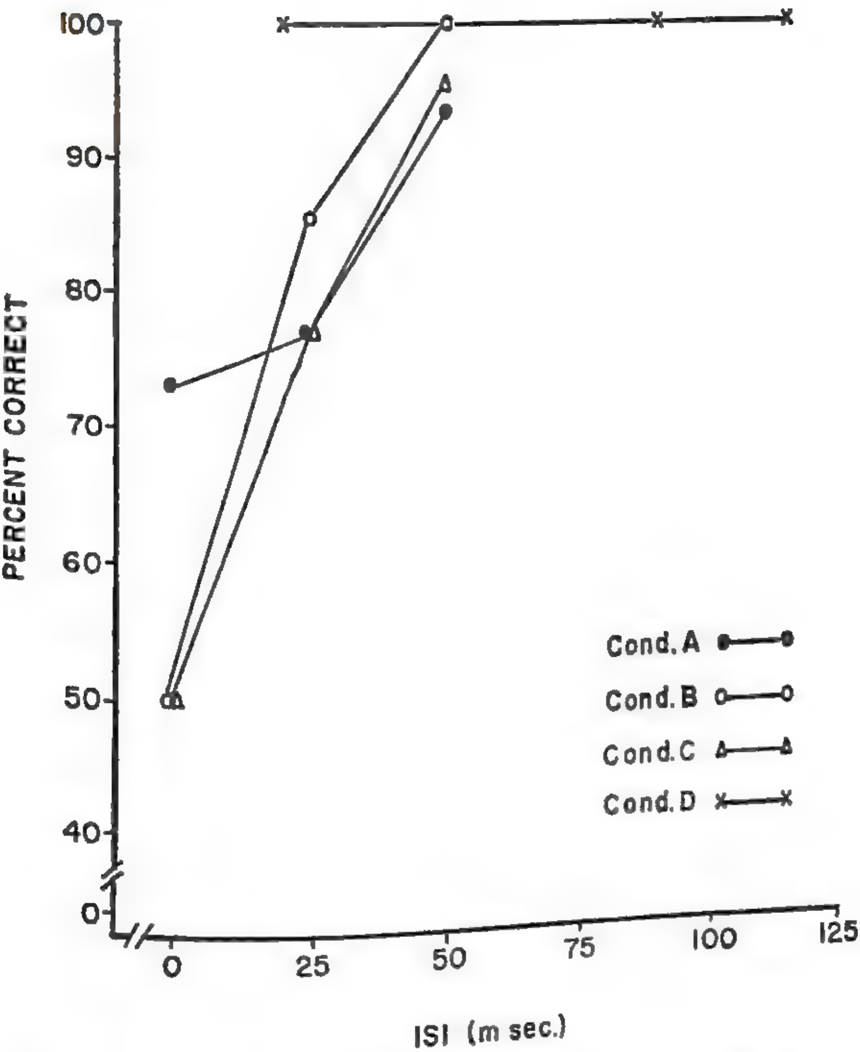


Fig. 2. Percentage Correct Forced-Choice Identifications as a Function of Interstimulus Interval (ISI) between Target and Masking Flashes: Experiment I

disinhibition, showed no less masking than was obtained under our Condition A. Some support for disinhibition effect might be deduced from the fact that Condition C did not show any *more* masking at the 25- and 50-msec. ISI than did Condition A. The argument would be that in terms of luminance summation, Condition C should have provided a greater masking at these ISI values than was obtained for the single masking stimulus in Condition A. The

fact that masking was not greater with two masks could be attributed to some sort of disinhibition effect.

If this argument were to be pursued, however, the temporal separation between the two masking flashes—separation which Robinson considered essential—would have to be discarded. Condition B, under which F1 and F2 occurred simultaneously, also showed no greater masking at ISIs of 25 and 50 msec. than was obtained in Condition A, where only the F1 masking stimulus was employed.

The obtained results are not what would be expected from a simple application of a luminance summation–contrast reduction explanation. In accordance with Bloch's law, the three stimuli would be expected to sum over the ISI values studied. Complete summation should yield the same effective stimulation as is obtained when all three stimuli (target, F1, and F2) are presented simultaneously. This simultaneous presentation would be a circle of light having a center disk of 3.6 mL. surrounded by a ring of 2.4 mL. which in turn is surrounded by a ring of 1.2 mL. Failure of Conditions B and C to remain at the chance level at 25-msec. ISI, as they were when F1 and F2 were presented simultaneously, clearly indicates that more than a simple summation is involved. More than simple summation is also indicated by the data from Condition D. This condition was essentially the same as Condition A except that F2, the larger disk masking stimulus, was employed instead of F1. When F2 occurs at 20-msec. ISI following the target, performance ought to be essentially the same as when F1 occurs 25 msec. following the target in Condition A. However, as is seen in Fig. 2, no masking was obtained at any ISI value when the F2 masking stimulus only was employed. This result suggests that the larger size of this masking disk may have been the relevant variable.

If the *Ss'* introspective reports, supplemented by *E*'s own observations of the stimuli under the different conditions, are examined, it is seen that these tasks provide a variety of rather complex cues which can be used in detecting the occurrence of the target. One of the most obvious cues is an apparent-movement effect that is obtained when all three stimuli are presented in certain temporal relationships. This apparent movement occurs between the target and F1 and the target and F1 and F2 in Conditions A, B, and C, with the nonsimultaneous ISI values. One phenomenally sees a light that expands from the center outward. Kahneman has also reported this effect, and it is to be noted that this apparent movement can provide a reliable cue as to whether or not the target disk has been

presented.[10] Under Condition D, where only the target and the larger mask (four times the diameter of the target) are presented, a different kind of cue is obtained. This cue appears to be based on the phenomenally bright and dark bands which are observed alternating toward the center of a lighted disk (Mach; Ratliff).[11] If only F2 is presented, this phenomenal darkening is noted; but if the target has occurred prior to the presentation of F2, the boundary of the target is enhanced while the darker bands just outside the target boundary appear even darker.

Experiment II

The presence of cues such as apparent movement and Mach bands in the larger disks served to confound any possible disinhibition effects that might have been present in the above experiment. The occurrence of these cues also indicates that the particular experimental task chosen by Robinson is probably inadequate for testing for disinhibition. Therefore, Experiment II was designed to eliminate these confounding effects. The target stimuli were the capital letters A, T, and U, and the two masking stimuli were circular lighted fields of equal size. The possibilities of apparent-movement cues were greatly reduced by having both masking fields of the same size and by having the target stimuli of incompatible shape with the masking stimuli. In addition, the use of the capital letters A, T, and U as targets permitted an identification criterion which should be more sensitive than a judgment of presence or absence in determining possible disinhibition effects.

Besides looking for possible disinhibition effects, Experiment II was concerned with another equally important question. This question was whether a time-intensity reciprocity (Bloch's law) holds for masking flashes. If a second masking flash is presented at an ISI too far removed from the termination of the target to effect target identifiability, does this second masking flash show summation with an earlier masking flash? Time-intensity reciprocity leads to the prediction that two masking flashes presented within the critical duration should sum to produce essentially the same masking effect that a single masking flash of the same total energy would produce when presented at the ISI of the first of the two masking flashes.

Apparatus. A 10-channel tachistoscope was again employed for the stimulus presentation. The target stimuli were the capital letters A, T, and U, white on

[10] Kahneman, *loc. cit.*
[11] F. Ratliff, *Mach Bands*, 1965.

black. They were presented transilluminated with a contrast in excess of 98%. The letters subtended 30 min. of visual angle in height and were on a background of 0.6-mL. intensity. The two masking flashes F1 and F2 were circular in shape, 2° in diameter, and 1.2 mL. in intensity. Contrast between the masks and the background was in excess of 99% and both masks were positioned to center on the target forms.

Subjects. Two graduate students (women) and a faculty member (a man) served as *S*s. All had normal or corrected-to-normal vision and had had previous experience in perceptual-recognition experiments.

Procedure. Before beginning the experiment proper, each *S* served in a practice session during which an exposure duration for the target forms was determined such that forced-choice identification accuracy was approximately 70% when the target form was followed by the F1 stimulus at an ISI of 20 msec. This criterion was achieved with an exposure duration of 2 msec. for two *S*s and 3 msec. for the third. Each *S* then served five experimental sessions, during each of which, 24 trials were administered under each of four different treatment conditions. Under Condition 1, one of the three target letters was presented and was followed, 20 msec. after termination, by a 10-msec. flash of the F1 stimulus. In Condition 2, F1 was not presented. Instead F2 was presented for 10 msec. at a 50-msec. ISI following the target. Condition 3 was essentially the same as Condition 1 except that F1 and F2 were presented simultaneously for 10 msec. at the 20-msec. ISI. In Condition 4, the stimulation sequence was: target letter, 20-msec. ISI, F1 for 10 msec., 20-msec. ISI, F2 for 10 msec. The three letters were presented in random order within the restriction that no letter appear more than three times in a row. In all conditions the adapting field and the ISI were dark except for a faint fixation cross subtending 20 min. of angle and appearing below the letters. Each *S* made 120 judgments under each of the four experimental conditions, which were presented in a counterbalanced order within and across *S*s and sessions. Under all conditions *S*s were required to respond with a choice among the three possible stimuli.

Since the target and the masking flashes in all four conditions occurred within a total duration that is less than the typical critical duration reported for these luminances, it was considered desirable to in fact determine whether time-intensity reciprocity could be obtained under the conditions of the present experiment. Accordingly, a final experimental session was devoted to a two-interval, forced-choice, brightness-judgment task. The standard stimulus here was a circular patch of light 10 min. of angle in diameter, with 2.4-mL. luminance and 10-msec. duration. This standard was compared with (*a*) two disks of the same diameter and 1.2-mL. intensity, each presented for 10 msec. with a 0.1-msec. ISI between offset of the first and onset of the second (essentially consecutive), (*b*) two disks of light of the same diameter, 1.2-mL. intensity, and 10-msec. duration, but separated by an ISI of 30 msec., and (*c*) two disks of light, at 10-msec. duration and 1.2-mL. intensity, presented with a 60-msec. ISI. Twenty-four forced-choice judgments were obtained from each *S* for each of the three conditions. Either the standard stimulus or a comparison stimulus occurred in the first interval 50% of the time and on a random basis, followed 0.5 sec. later by the other stimulus. The *S* was required to choose which interval contained the brighter stimulus.

Results. Table I shows the average percentage correct letter identifications under the four experimental conditions. The target forms were identified with 70.3% accuracy under Condition 1, where only a single masking flash was presented and at an ISI of 20 msec. In Condition 2, where only the F2 masking stimulus was presented and at an ISI of 50 msec., identification performance rose to 95%, a performance level essentially the same as that obtained when no masking flashes were presented. Since F1 and F2 were identical in luminance and duration, the improved performance under Condition 2 is attributable to the increased ISI between form termination and onset of the mask. The poorest performance, 44.4%, was obtained under Condition 3, where both F1 and F2 occurred simultaneously at an ISI of 20 msec. This condition is equivalent to a single masking flash at twice the luminance that obtained in Condition 1. Of greatest interest is performance under Condition 4. Here F1 occurred at an ISI of 20 msec. and F2 occurred at an ISI of 50 msec. following termination of the target form. As Table I shows, performance under this condition was midway between that obtained under Condition 3 and Condition 1. Statistical analyses performed on these data show that for all three *Ss*, performance under any of the four experimental conditions differed significantly from the performance under any of the other experimental conditions ($p < 0.05$).

From these data it is apparent that some summation was occurring between the F1 and the F2 masking flashes in Condition 4, since performance was appreciably poorer than that obtained when only the F1 mask was employed at 20-msec. ISI or only the F2 mask at 50-msec. ISI. However, when this performance is compared with that under Condition 3, it is apparent that summation was not total, for performance was 44.4% when both masks were presented simultaneously at 20-msec. ISI as opposed to the 57.2% accuracy obtained in Condition 4. Failure to obtain complete summation in terms of masking effects for F1 and F2 in Condition 4 is at variance with what would be anticipated from prior studies on the temporal

TABLE I

MEAN PERCENTAGE CORRECT LETTER IDENTIFICATIONS AND PERFORMANCE RANGE

Condition	Percentage correct	Range
1	70.28	61.67–75.83
2	95.00	92.50–98.33
3	44.44	40.00–50.83
4	57.22	45.00–67.50

summation of light pulses.[12] On the other hand, this result is consistent with recent findings that there is no critical duration over which complete time-intensity reciprocity holds for intermittent light stimulation on a form-identification task.[13]

A brightness criterion was used to examine the extent to which the F1 and F2 masking stimuli sum over temporal separations. In the final experimental session, when *Ss* were presented with the two-interval forced-choice between a 20-msec. flash of 1.2-mL. luminance and a 10-msec. flash of 2.4-mL. luminance (equivalent to the F1 and F2 masks occurring simultaneously), the *Ss* reported that the 20-msec. flash was brighter 48% of the time. This was not a significant nor appreciable difference from the expected 50% for equal brightness. However, when there was an interruption in the form of an ISI in the presentation of the energy, reciprocity broke down. The 10-msec. continuous flash of 2.4 mL. was judged brighter 64% of the time when compared with two 10-msec. flashes of the same luminance separated by 30-msec. ISI. When the two pulses were separated by 60-msec. ISI, the 10-msec. standard was judged brighter 88% of the time.

Discussion

In neither experiment were we able to confirm Robinson's "disinhibition effect." In Experiment I, which closely duplicated the conditions of Robinson's experiment, masking was unaffected when a second masking flash followed the first at ISI values where Robinson had obtained disinhibition. The *Ss'* reports of their phenomenal experiences in viewing the set of stimuli employed by Robinson suggest that these particular stimuli provide a very complex perceptual experience containing a number of possible cues for discriminatory responses. A small target circle followed by two masking circles of progressively larger size gives rise not only to apparent-movement effects but also to effects that appear to be related to Mach bands. An experienced *S* has several possible cues available to him in making discriminatory responses as to the presence or absence of the small target circle. With these various cues available, it is impossible to disentangle any possible disinhibition effects from the other effects that are occurring.

These extraneous cues also would account for the failure of the data in Experiment I to reveal a complete luminance-summation effect. In Experiment II, where apparent movement and Mach band

[12] Lichtenstein and Boucher, *loc. cit.*; Davy, *loc. cit.*
[13] D. L. Schurman, C. W. Eriksen, and J. Rohrbaugh, Masking phenomena and time-intensity reciprocity for form, *J. exp. Psychol.*, 78, 1968, 310–317.

effects were eliminated by using letters as the target stimuli and having the two masking circles of equal size, the results are consistent with expected temporal luminance-summation phenomena. However, this energy summation was not total as might have been anticipated from prior research.[14] The introduction of a 20-msec. dark period (ISI) between the two masking flashes resulted in less than total summation even though the 20-msec. interval between flashes, and the total interval over which the masking flashes were distributed (40 msec.), was less than the generally accepted critical duration.

The failure to obtain complete summation over a total time of 40 msec. is apparently not attributable to the masking criterion that was employed. Control observations indicate that a 30-msec. ISI between two masking flashes is sufficient to result in the two flashes appearing less bright than a single flash containing the same total energy. This finding suggests that the generality of Bloch's law to pulsed or intermittent light stimulation is limited. Recent work has shown that Bloch's law is not applicable to tasks involving form identification.[15] The present data, although not very extensive, suggest limitations of this law in terms of an apparent-brightness criterion.

Of particular interest in this study was a temporal bridging effect obtained in Condition 4 of Experiment II. It was found in Condition 2 that when the F2 mask alone was presented at an ISI of 50 msec. following the target, little or no masking effect was obtained. Identification accuracy was 95%, essentially the same as that obtained when no masking stimuli were presented. When the F1 masking stimulus was interpolated between termination of the target and the F2 mask, there was an appreciable increment in the amount of target masking over that obtained by the F1 mask alone. Thus the F2 mask occurring at a temporal ISI that by itself was too great to effect target identifiability, did reduce the target identifiability when mediated by an interpolated F1 mask. How this effect is mediated, and over how long an ISI value it obtains, was not answered in the present experiment but is a subject for further research.

Summary

Robinson has reported disinhibition of visual masking when a second masking disk follows a first masking disk at a certain in-

[14] Lichtenstein and Boucher, *loc. cit.*; Davy, *loc. cit.*
[15] Schurman, Eriksen, and Rohrbaugh, *loc. cit.*

terstimulus interval (ISI). The first of the present two experiments replicated Robinson's experiment, except that a temporal forced-choice indicator methodology was employed. Not only did the results fail to confirm Robinson's findings, but the task and procedure were found to contain a number of cues such as Mach bands and apparent movement that would completely confound any possible disinhibition effects. The second experiment was designed to eliminate these confounding cues and also to determine whether a masking flash temporally too far removed from the target to mask by itself would sum with an intervening masking flash to produce increased masking. This experiment failed to confirm disinhibition effects but did find that a second masking flash occurring at a temporal ISI that by itself was too great to effect target identifiability, did reduce the target identifiability when mediated by an interpolated first masking flash.

THE EFFECT OF EXPRESSIVE VERBAL REINFORCEMENTS ON INCIDENTAL LEARNING BY MODELS AND OBSERVERS

By SEYMOUR M. BERGER and SUSAN W. ELLSBURY,
Indiana University

This study was an outgrowth of a previous investigation by Berger[1] and some suggestive findings by Rosenthal, Kohn, Greenfield, and Carota.[2] Berger's paper provided some evidence that the incidental learning of an observer can be influenced by the verbal reinforcements presented to a model, *i.e.* by vicarious reinforcement. The verbal-reinforcement conditions there reported required that *E* say "right" or "wrong" to a model's responses, which were preceded by incidental cues. The verbal reinforcers were consistently associated with particular incidental cues during learning. When tested to determine which incidental cues they recalled, the observers tended to recall more of the cues associated with the model's 'right' responses than 'wrong' responses. This effect was present in each of three experiments run, although it was statistically reliable only in the first. In two of the three experiments, the models also recalled more incidental cues associated with "right" than "wrong," but in each case the effect was small and not statistically reliable.

More recently, Rosenthal, Kohn, Greenfield, and Carota reported a low but positive correlation between verbal-conditioning scores and the expressiveness with which *E* announced verbal reinforcers.[3] This finding suggested to the present authors that Berger's results may have been influenced by the expressiveness with

* Received for publication January 30, 1969. The research reported was partially supported by National Science Foundation Research Grant GS-1702 to the senior author. The paper is based in part on Susan W. Ellsbury's undergraduate honors dissertation. The authors wish to express their appreciation to Margaret Peterson for the use of her laboratory facilities and to Gabriel P. Frommer for his helpful comments and suggestions.

[1] S. M. Berger, Incidental learning through vicarious reinforcement, *Psychol. Rep.*, 9, 1961, 477–491.

[2] R. Rosenthal, P. Kohn, Patricia M. Greenfield, and N. Carota, Data desirability, experimenter expectancy, and the results of psychological research, *J. pers. soc. Psychol.*, 3, 1966, 20–27.

[3] Rosenthal, Kohn, Greenfield, and Carota, *loc. cit.*

which the verbal reinforcers were presented in his three studies. More specifically, it seemed reasonable that E might have been more enthusiastic in the initial study and less enthusiastic in his delivery of reinforcement in the later studies, when the procedure became more routine. Furthermore, there seemed to be a theoretical basis for assuming that the emotional tone of E's voice might have influenced the observers' learning: several studies have shown that an observer's learning can be affected by the emotional state of someone else.[4] If the observers in Berger's three studies had experienced *vicarious pleasure* when their model was 'right' and *vicarious displeasure* when their model was 'wrong,' one might expect these vicarious emotional experiences to have differentially influenced the observers' recall in favor of cues associated with pleasant experiences. Similarly, the models' recall might have been influenced by the emotional tone of the reinforcements delivered directly to them.

This analysis forms the background for the present study, which employed basically the same experimental task used by Berger.[5] In this study, however, one group of models and observers heard E deliver the reinforcements in an enthusiastic fashion while a second group heard the reinforcements delivered in a nonenthusiastic manner. The hypothesis tested was that observers and models in the Enthusiastic condition would show greater differential recall of incidental cues associated with "right" and "wrong" than observers and models in the Nonenthusiastic condition. That is, if the hypothesis regarding vicarious emotional reinforcement were correct, then Ss in the Enthusiastic condition would recall more incidental cues associated with "right" and fewer cues associated with "wrong" than Ss in the Nonenthusiastic condition.

Method

Subjects. The Ss were 120 students at Indiana University who were enrolled in an introductory psychology course; their participation in the experiment fulfilled a course requirement. Sixty Ss acted as models, and 60 Ss acted as observers.

Apparatus. A Lafayette memory drum was used to show 16 nonsense syllables, typed in capital letters, to both models and observers. The drum rotated every 2 sec. A blank space appeared between each nonsense syllable; six blank spaces appeared between each of the trials. E sat across the table

[4] A. Bandura and T. L. Rosenthal, Vicarious classical conditioning as a function of arousal level, *J. pers. soc. Psychol.*, 3, 1966, 54–62; S. M. Berger, Conditioning through vicarious instigation, *Psychol. Rev.*, 69, 1962, 450–466.

[5] Berger, *op. cit., Psychol. Rep.*, 9, 1961, 477–491.

from the Ss. A screen around the drum obscured the recording of the responses.

Procedure. Two Ss were brought into the experimental room and asked to take seats on one side of the table facing the memory drum. The S on the left was told he was the 'subject' in the experiment. Both Ss were informed that their assignments were based solely upon the seats they had chosen: the S seated to the E's left was the *model;* the other S was the *observer.* These instructions were given:

Perhaps you have heard of extrasensory perception, or mental telepathy. Some investigators believe that they have clearly demonstrated that people can receive thought messages from someone else. Other investigators have not been able to demonstrate extrasensory perception and are less convinced. The purpose of this experiment is to study extrasensory perception by a method that is somewhat different from the ones used before.

I have prepared a list of nonsense words and numbers. The subject in this experiment is to try to judge what number I have paired with each of the nonsense words. If extrasensory perception does exist, the subject should be able to judge the paired numbers better than would be expected by chance. If extrasensory perception does not exist, the subject will guess correctly only on some occasions, which is expected by chance. The numbers I have chosen vary between 1 and 10 (1 through 10), so you may choose from 10 numbers in each case.

[To model:] The purpose of pairing the numbers with nonsense words is to standardize the experience of the subject at the time the thought message is sent. In order to assure that the subject has this standardized experience, I want you to watch the rotating drum, and as the nonsense word appears read the nonsense word aloud. Then make an immediate judgment as to what you think the paired number might be. You must give a judgment on *every* trial. Thinking about your past judgments will only interfere with the extrasensory processes.

In order to assure that I make no mistake in recording your judgments on my data sheet, I have found it convenient if you would announce the nonsense word you are looking at first and then make your judgment. So, for example, if the word is "QXR," and your judgment is "5," simply announce "Q-X-R-5," and I will record it. Notice that you spell the letters of the word, do not try to pronounce the word.

I have a list of 16 words, and this list will be presented a number of times. But the numbers assigned to words will be changed in random fashion from presentation to presentation. Thus, a number which is correct on one presentation will probably not be correct on the next presentation.

This experiment is designed to move along quickly so as to minimize interference effects. As we go along I will try to tell you if your judgments are right or wrong. I have found, however, that I occasionally do not have time to keep up with the experiment and with recording the data. So rather than interfere with your attention to the next word, I will occasionally just not say anything about whether you are right or wrong. When the experiment is over, however, we will go over your judgments together and you will be able to see how well you did.

Are there any questions? Remember, please do not debate—give the first number that comes to mind. And once the experiment begins, please do not talk to each other, since we do not want the subject to be distracted. There will be one practice trial and six experimental trials.

[To observer:] Because of my interaction with the subject, I must concentrate on the number which I am thinking; therefore, I will not have time to operate the machine. All you have to do is turn the switch on and off at the appropriate times. [The observer was shown how the machine operated.]

The model tried to guess the number E had supposedly paired with each of the given nonsense syllables. The observer's task was to operate the drum, turning it off and on between trials; he was to watch the presentation of syl-

lables also, although not required to read them aloud as was the model. Verbal reinforcement was given by a prearranged schedule in accordance with the syllables, not the number responses.

The 16 nonsense syllables used were of low association value based on Glaze's findings.[6] The average association value for the *critical* syllables was 7% with the range from 0% to 20%. The *S*s received one practice trial, in which the responses were not reinforced, and six experimental trials. The 16 nonsense syllables were presented in the same order on each trial. In order to weaken the effect of serial position on recall, the first two and last two syllables were always reinforced as 'wrong' and were not used in the analysis. It was hoped that this procedure also would enhance the credibility of the model's performance, since the model would make incorrect guesses most of the time. The middle 12 syllables were considered the critical syllables in the experiment: four were reinforced as 'right,' four were reinforced as 'wrong,' and four were not reinforced. Each pair of *S*s received only one of the schedules of reinforcement during the trials. Through counterbalancing of the reinforcement schedules for syllables over *S*s, each syllable received an equal number of "rights," "wrongs," and no responses by the end of the experiment. Six reinforcement schedules were used with 10 pairs of *S*s receiving each arrangement. The assignment of *S*s was by random selection.

*S*s were alternately assigned to either the *Enthusiastic condition* or the *Nonenthusiastic condition*. In the Enthusiastic condition, *E* read the instructions with voice inflections; in the Nonenthusiastic condition, *E* read them in a bland tone. In the Enthusiastic condition, 'right' and 'wrong' reinforcements were given with appropriate positive and negative emotional intonations. In addition, the experimenter said either "good," "fine," or "great" in an excited fashion following the reinforcement "right." In the Nonenthusiastic condition, "right" or "wrong" was given in an unemotional tone.

After six trials, both model and observer were given a questionnaire. The *S*s were told to rate each question honestly and that *E* would not see their answers. The questionnaires were folded and put into a covered box. Actually only one of the five questions was specially relevant to the experiment; the rest were fillers to help disguise the true purpose of the questionnaire, which was to obtain a validation of *E*'s attempt to act enthusiastic. The relevant question was: "Did the experimenter seem enthusiastic about the experiment?" Below this question was a 9-point scale with "very enthusiastic" at one end and "very nonenthusiastic" at the other. Next, the *S*s each were given a blank sheet of paper and asked to write down, independently, all of the syllables they could recall from the list. A maximum of 5 min. was given for this task. When the experiment was completed, the *S*s were told its purpose.

Results

The analysis of variance of correct recall scores for both types of *S* (model and observer) revealed two main effects. *S*s in the Enthusiastic condition correctly recalled more nonsense syllables than did *S*s in the Nonenthusiastic condition ($F = 9.57$, $df = 1/116$,

[6] J. A. Glaze, The association value of nonsense syllables, *J. gen. Psychol.*, 35, 1928, 255–269.

$p < .01$). The mean recall in the Enthusiastic condition was 2.57, while the mean for the Nonenthusiastic condition was 1.63. Type of reinforcement also produced a significant overall effect ($F = 5.11$, $df = 2/232$, $p < .01$). None of the interactions was significant. A comparison of the effects of three types of verbal reinforcements ("right," "wrong," and no response) showed that more of the nonsense syllables associated with "right" were recalled than those associated with either "wrong" or no response (Duncan's multiple-range test, $\alpha = .05$). The mean recall scores for models and observers are reported in Table I. The lack of a main effect for subjects or of significant interactions demonstrates that the effects reported above for conditions and type of reinforcement are characteristic of the recall scores for both models and observers.

In order to verify independently the effectiveness of the variation of *E*'s enthusiasm, the *S*s' questionnaire responses were examined by analysis of variance. This analysis dealt with the mean ratings given by both types of *S*s to the question, "Did the experimenter seem enthusiastic about the experiment?" Only a significant main effect for conditions was found ($F = 27.47$, $df = 1/116$, $p < .01$). The experimenter was rated as more enthusiastic in the Enthusiastic condition than in the Nonenthusiastic condition: the means were 7.92 and 6.57, respectively. In general, *S*s tended to give favorable ratings, using the 'enthusiastic' side of the scale almost exclusively.

Discussion

The results did not support the hypothesis which prompted the study. Although *E*'s enthusiasm did facilitate learning of the incidental materials, the effect was *general* rather than *specific*. The vicarious-experience hypothesis predicted that *S*s exposed to an enthusiastic *E* who announced the verbal reinforcements in an excited and emotional tone would learn more syllables associated with "right" and fewer syllables associated with "wrong" than *S*s

TABLE I

Mean Recall Scores

*S*s	Enthusiastic condition				Nonenthusiastic condition			
	Right	Wrong	No response	Total	Right	Wrong	No response	Total
Models	.97	.90	.83	2.70	.70	.40	.47	1.57
Observers	1.00	.73	.70	2.43	.83	.40	.47	1.70

who had a nonenthusiastic *E*. Models and observers did learn more of the syllables associated with "right" than with either "wrong" or no response. The effect of enthusiasm, however, was simply to increase the total number of items learned regardless of the type of reinforcement used.

Several investigators have suggested that the emotional reactions of others may produce specific empathetic or vicarious experiences in an observer,[7] and some have suggested that such experiences can reinforce learning. The present study did not provide evidence that the emotional reactions of others have specific reinforcing effects. This same conclusion was drawn from another study which attempted to test for specific effects of a model's emotional reactions in observer learning.[8] In that study, an observer was exposed to a model who was reinforced either 100% of the time or just 25% of the time. Within each of these reinforcement conditions, half of the observers were exposed to a model who reacted with pleasure when her responses were correct and with displeasure when she was incorrect; the other half of the observers watched a nonemotional model under the same conditions. The results of that experiment showed that observers who watched a partially reinforced model had greater resistance to extinction than observers who watched a continuously reinforced model; the effect of the model's emotional reactions was generally to increase the observer's resistance to extinction regardless of the reinforcement condition. Therefore, in addition to the present study, there is evidence for the interpretation that emotional reactions of others have general rather than specific consequences for observer learning.

In order to provide further support for this interpretation, some additional analyses of the data were performed. If *S*s in the Enthusiastic condition were more highly motivated than *S*s in the Nonenthusiastic condition, this difference in motivation might be manifested in the *total* number of responses given on the recall test. The more motivated *S*s should give more total responses (*i.e.* correct and incorrect responses) on this test than the less motivated

[7] J. Aronfreed, The origins of altruistic and sympathetic behavior, a paper read at the APA meeting, New York, 1966; A. Bandura and R. H. Walters. *Social Learning and Personality Development*, 1963, 81–84; Berger, *op. cit.*, *Psychol. Rep.*, 9, 1961, 477–491; Berger, *op. cit.*, *Psychol. Rev.*, 69, 1962, 450–466; J. Kagan, The concept of identification, *Psychol. Rev.*, 65, 1958, 296–305; O. H. Mowrer, *Learning Theory and the Symbolic Processes*, 1960, 115–116.

[8] S. M. Berger and Sandra L. Johansson, The effect of a model's expressed emotions on an observer's resistance to extinction, *J. pers. soc. Psychol.*, 10 1968, 53–58.

Ss. An analysis of variance of the total responses given by each type of *S* in the two conditions revealed a significant main effect only for conditions ($F = 5.92$, $df = 1/116$, $p < .05$). *Ss* in the Enthusiastic condition gave more total responses on the recall test than *Ss* in the Nonenthusiastic condition; the means were 6.08 and 5.05, respectively. The effect was the same for both models and observers.

An analysis of variance was also performed on the ratings given by *Ss* to one of the questionnaire's filler items—an item asking them to rate how interesting they found the experiment. Higher ratings of interest in the experiment were expected from the more motivated *Ss* than the less motivated *Ss*. In this case, both main effects were significant: for conditions, $F = 9.35$, $df = 1/116$, $p < .01$; and for subjects, $F = 8.84$, $df = 1/116$, $p < .01$. The interaction was not significant. *Ss* in the Enthusiastic condition reported more interest in the experiment than did *Ss* in the Nonenthusiastic condition; the mean ratings were 7.03 and 5.83, respectively. With regard to the main effect for subjects, the models indicated greater interest in the experiment (mean = 7.01) than did the observers (mean = 5.85). That is, in both conditions, greater interest in the experiment was reported by the models than by the observers, but both models and observers in the Enthusiastic condition reported greater interest in the experiment than did the models and observers in the Nonenthusiastic condition.

Just how the assumed change in general motivation affects incidental learning was not clarified by the present experiment. One effect of *E*'s enthusiasm may be to increase *S*'s interest in the task and his attention to the materials presented. Another possible explanation is based on the assumption that moderate increases in general activation enhance performance. Perhaps exposure to someone else's emotional state increases observer activation and, as a consequence, enhances observer learning.[9] In the present study, *both* the model and the observer were in fact *observers* of *E*'s emotional state; thus an increase in activation may have been produced in both *Ss* in the Enthusiastic condition, with an attendant increase in learning.

This study replicated the finding that observers' recall of incidental materials was differentially affected by the reinforcements

[9] S. M. Berger, Vicarious aspects of matched-dependent behavior, in E. C. Simmel, R. A. Hoppe, and G. A. Milton (eds.), *Social Facilitation and Imitative Behavior*, 1968, 169–185.

given to a model.[10] In the present study, however, the models' recall also was differentially affected by the reinforcements, an effect which was not clearly evident in the earlier study. Furthermore, in this study no differences were found in the recall scores of the model and observer, while in the earlier study observers tended to recall more items than the models. In the earlier study the observers read the syllables to the model, while in the present study the observer's task was merely to operate the memory drum. This difference in the focus of the observer's task may have reduced his learning advantage over the model.

The implications of this study for laboratory studies involving verbal reinforcement are quite clear. The effectiveness of these reinforcements depends in part on the enthusiasm shown by *E*. If *E* is engaged in a relatively long experiment, with a large number of *S*s, his initial enthusiasm may wane over an extended period, with a consequent reduction in reinforcement effectiveness. The same implications apply to learning in natural settings as well. The teacher or parent, for example, who puts some emotional force behind verbal reinforcements is likely to exert greater influence. It must be noted, however, that such influence may not be in the intended direction. In accordance with the findings of the present study, expressive verbal reinforcements not only enhance learning of the specific responses they follow but also enhance learning of other responses elicited by the situation—even of those responses which were negatively reinforced or not specifically reinforced at all!

It is also clear that *S*s' recall is influenced by the verbal reinforcer "right"; verbal reinforcers do produce specific effects in incidental learning. *S*s in this experiment may have entertained the hypothesis that the nonsense syllables contained some hidden clue to the correct number response. Consequently, when the model was 'right' both model and observer may have reviewed or covertly rehearsed the syllable-number sequences which were 'right' to determine if some pattern emerged. Such extra rehearsal could increase learning of the syllables associated with the reinforcer "right."

It is possible that the procedures used in this study tended to separate the informational and motivational components of verbal reinforcers. The informational component of verbal reinforcement may have been represented by the reinforcers "right" and "wrong." The motivational component may have been carried in the expres-

[10] Berger, *op. cit., Psychol. Rep.*, 9, 1961, 477–491.

sive manner in which the reinforcements were presented in the experiment. Consequently, the procedures used in this study may have general implications for the analysis of the effectiveness of verbal reinforcement in terms of informational and motivational factors.

Summary

The purpose of this study was to determine whether a previously demonstrated effect of vicarious reinforcement on observers' incidental learning could be enhanced if the experimenter announced "right" and "wrong" in an expressive-emotional manner. Half of the *S*s heard the reinforcements delivered in an enthusiastic fashion, while the other half heard the reinforcements given in a nonenthusiastic manner. Both models and observers recalled more of the incidental cues associated with the reinforcer "right" than either with "wrong" or with no response from *E*, but the effect of *E*'s enthusiasm was to enhance recall of all syllables, regardless of the type of reinforcement. The results did not support a vicarious-experience hypothesis. Some alternative hypotheses were suggested.

EQUIDISTANCE EFFECTS IN VISUAL FIELDS

By WALTER C. GOGEL,
University of California, Santa Barbara

In a visual field with a surface (floor) located below the eyes of the *O* and extending in depth, the parts of the floor that are more distant also are higher in the field of view. It follows that an object suspended above such a floor tends to be nearer in its visual direction from *O* to the more distant portions of the floor. This is illustrated in Fig. 1, a perspective line drawing which represents *O*'s view of a visual alley and a suspended object (A). It will be noted that the most distant parts of the visual alley are the parts directionally closest to Object A. As Fig. 1 suggests, when there are no cues to locate an object like A in depth, it appears to be near the back surface of the alley regardless of the distance at which it might be physically located in the alley. If, however, an object like A is lowered in the visual field, its base can directionally intersect the floor of the visual field. When this occurs, A is perceived at a distance in the alley where a line from the eye of *O* through the base of Object A intersects the alley floor. This phenomenon, described by Gibson as the cue of optical contact,[1] has been used to determine apparent size and apparent distance,[2] and has been investigated as a function of perceptually different background slopes.[3]

The cue of optical contact can be considered as a special case of a perceptual factor known as the equidistance tendency:[4] the tendency, in the absence of effective distance cues, for objects (or parts of objects) to appear equidistant from *O*, with a strength inversely related to the directional (angular) separation of the objects. The tendency is assumed to exist with respect to all objects in the visual field, from which assumption it follows that the

* Received for publication January 8, 1968.

[1] J. J. Gibson, *The Perception of the Visual World*, 1950.

[2] R. P. McDonald and Patricia T. O'Hara, Size-distance invariance and perceptual constancy, this JOURNAL, 77, 1964, 276–280.

[3] W. Epstein, Perceived depth as a function of relative height under three background conditions, *J. exp. Psychol.*, 72, 1966, 335–338.

[4] W. C. Gogel, Equidistance tendency and its consequences, *Psychol. Bull.*, 62, 1964, 158.

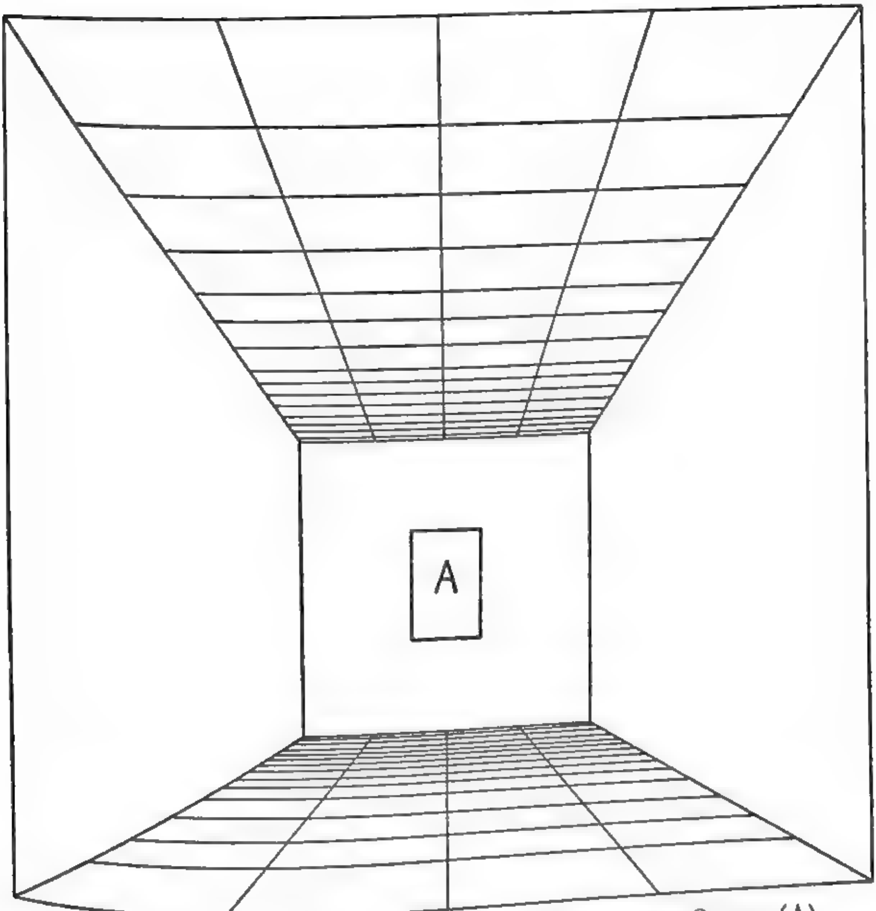


FIG. 1. *O*'S VIEW OF A VISUAL ALLEY AND A SUSPENDED OBJECT (A)

apparent distance of an object with few or no effective cues to distance will be determined as a resultant of the combined effect of all the equidistance tendencies between it and other parts of the visual field.

Under the conditions illustrated in Fig. 1, therefore, the strength of the equidistance tendency would be greatest between Object A and the most distant parts of the alley, since these most distant parts are most adjacent in direction to Object A. Consequently, Object A, under these conditions, would appear at a distance approximating the most distant parts of the visual alley. Object A would not necessarily appear to be located *at* the distance of the back wall,[5] particularly if the back wall were a black area that did

[5] W. C. Gogel, Perception of depth from binocular disparity, *J. exp. Psychol.*, 67, 1964, 379–386.

not appear as a surface. But since the strength of the equidistance tendency is greatest between objects that are in the same direction, it would be expected that if Object A were to make optical contact with some portion of the floor (or back wall) of the alley, then A would tend to appear at or near the depth of that portion of the floor (or wall). Furthermore, it would be expected from the equidistance tendency that under circumstances in which all portions of Object A were in optical contact with the floor, Object A would sometimes appear to be stretched along the floor. This would occur when cues for depth were not available to indicate to *O* that all parts of the object were at the same distance.

The assertion that optical contact is a special case of the equidistance tendency is supported in that optical contact cannot explain the apparent depth of an object when no parts of the object are in optical contact with any part of the visual field.[6] The equidistance tendency can explain both this case and the case of perceived distance resulting from optical contact. The equidistance tendency also has implications for the not-infrequent use of visual fields like that shown in Fig. 1 to measure the perceived distance resulting from particular cues associated with experimental objects like A.[7] The use of a comparison field, particularly when it is presented simultaneously with the experimental object, is open to the objection that the comparison field itself produces equidistance effects with respect to the experimental object. Suppose, for example, in Fig. 1, that the capability of the cue of accommodation to localize Object A close to *O* were being evaluated. This localization would be opposed by the equidistance tendency making Object A appear distant from *O*, *i.e.* toward the far end of the alley. Thus, the equidistance tendency could modify the perceived distance that would have resulted from the cue system being investigated.

The present study had two objectives: (1) to demonstrate that an object can appear at the more distant portion of the field of view as a consequence of the equidistance tendency and (2) to demonstrate that equidistance effects between a comparison field and the experimental object presented simultaneously can distort the measurement of the effectiveness of a depth cue. Both of these objectives were investigated by comparing the perceived size and distance of an

[6] Gogel, *op. cit., Psychol. Bull.*, 62, 1964, 153–163.
[7] See W. C. Gogel, Size cue to visually perceived distance, *Psychol. Bull.*, 62, 1964, 217–235.

object when it was presented above the floor of a visual alley with the perceived size and distance of the same object presented in the dark. The cue system whose effect was modified by the equidistance tendency was that of familiar size.

Method

Apparatus. The familiar object used in the experiment was the 10 of spades from a deck of playing cards. A photographic transparency of the card was inserted between a source of light and a projection screen that was 10.6 ft. to the right of *O*.[8] The transparency was viewed monocularly by *O* (right eye only) by means of a partially transmitting, partially reflecting mirror which made the card appear to be in front of *O*. The visual angle of the transparency was equal to the visual angle of a normal-sized card at 5 ft. from *O*. The transparency had a luminance of 3 ftL. *O* was provided with a head-and-chin rest and looked through a viewing aperture located in a booth that was totally dark during the experiment.

The alley which *O* saw was 3 ft. wide and 16.7 ft. long, and had a floor composed of white and black checkboard squares (9 in. on a side) located 7 in. below his eye. The walls of this visual alley were formed by white cloth. Along the floor of the alley at a distance of 3, 6, 8, 12, and 16.3 ft. from *O* were a flashlight, a cigarette package, a pair of sunglasses, a soup can, and a butter carton, respectively. The average luminance of the alley was approximately the same as that of the playing card. The visual alley, like the card, was always viewed with the right eye only.

The simulated playing card (transparency) and the visual alley could be presented simultaneously or separately. When presented with the alley, the card was viewed against a black curtain at the end of the alley, with the bottom of the card 3 in. above the alley floor. When the card was presented alone, no other object or surface was visible.

Procedure. Forty students from an introductory course in psychology were used in the experiment. All *O*s had a visual acuity in their right eye of at least 20/20, both near and far, as measured with the Keystone orthoscope. Part of the instructions were:

This is an experiment on how people perceive objects. You will be asked to indicate what you perceive the size and distance of objects to be. When we ask you to indicate the distance of an object, we want you to tell us how far the object appears to be from your eyes. When we ask you to indicate the size of an object, we want you to tell us how wide (left-to-right extent) the object appears to be. You will indicate how far the object appears to be from your eyes, or how wide the object appears to be, by verbal reports which you will express in feet or inches or, if you wish, in some combination of feet and inches.

For all *O*s and all presentations, the distance report was completed prior to the size (width) report for each situation. Two kinds of experimental situations were used. In one situation (Alley condition), the visual alley and the simulated playing card were presented simultaneously. In the other situation (Reduced condition), only the playing card was visible in an otherwise totally dark

[8] W. C. Gogel and H. W. Mertens, A method of simulating objects moving in depth, *Percept. mot. Skills*, 23, 1966, 371–377.

visual field. Twenty *Os* were presented first the Reduced condition, then the Alley condition. The order of conditions was reversed for the 20 other *Os*. Judgments of the apparent distance and the apparent size of the card were made in both conditions, and then the alley was presented again, with the playing card absent. The purpose of this final presentation was to calibrate *O*'s estimates of apparent distance. Since adequate cues to perceived distance were present in the alley, it was hypothesized that errors in judging the distances of the objects (the flashlight, etc.) on the floor of the alley could be attributed to individual idiosyncrasies in mentally applying a foot rule to the estimation of distance. This last situation, in which the *Os* judged the distance of the five objects along the alley floor, was termed the calibration situation. A line of linear best fit, using the method of least squares, was applied to the results from the calibration data from each *O*. This permitted the estimated distances of the card that were obtained in the Alley and Reduced conditions to be converted to scores corrected for individual differences in using the technique of verbal estimation.[9] The scores resulting from this procedure were called calibrated scores.

Results

The results are summarized in Table I, with the original scores (verbal estimates) and calibrated scores (verbal estimates corrected for observer idiosyncrasies) shown in the upper and lower portions of the table, respectively. The effect of the equidistance tendency on the perceived distance, and thus on the perceived size, of the playing card is indicated by the differences between the results (average and median) from the Alley and Reduced conditions. For example, the first presentation of the card and alley to-

TABLE I

Reported Distance (D') and Reported Size (S')

	First presentations			Second presentations		
	Alley	Reduced	Difference	Alley	Reduced	Difference
Original scores						
Mean D' in ft.	7.5	3.8	3.7	9.7	4.5	5.2
Median D' in ft.	5.5	4.0	1.5	7.0	3.8	3.2
σ of D'	5.2	2.4		6.5	3.1	
Mean S' in in.	5.3	3.1	2.2	10.1	6.0	4.1
Median S' in in.	5.0	2.5	2.5	5.5	3.5	2.0
σ of S'	2.6	2.5		12.9	7.7	
Calibrated scores						
Mean D' in ft.	10.3	5.5	4.8	12.1	6.2	5.9
Median D' in ft.	8.2	5.2	3.0	12.4	6.3	6.1
σ of D'	5.3	4.4		2.9	2.6	

[9] W. C. Gogel, B. O. Hartman, and G. S. Harker, The retinal size of a familiar object as a determiner of apparent distance, *Psychol. Monogr.*, 71, 1957 (No. 13, Whole No. 442), 1–16.

gether resulted in an average report that the card was 7.5 ft. from O. When the card was first presented alone, however, it was reported on the average to be at a distance of 3.8 ft. Similarly, under these two conditions the average reported widths (sizes) were 5.3 and 3.1 in. respectively. These differences in distance (3.7 ft.) and in size (2.2 in.) were both significant at the .02 level ($t_{38} = 2.85$ and 2.68 respectively). Also, considering the second presentations, the 5.2-ft. difference between the average reported distances in the Alley and Reduced conditions was significant at the .01 level ($t_{38} = 3.19$). Although the mean differences between reported sizes in these two conditions were not significant, at the .05 level, the reported size differences (using the Mann-Whitney U test) were significant at the .01 level ($Z = 2.64$).

Table I shows that the average differences between the results from the Reduced and Alley conditions were consistent; *i.e.* the card presented in the Alley appeared to be more distant and larger than the card presented in the Reduced condition. The one exception was the average difference in size obtained between the first presentation of the Alley and the second presentation of the Reduced condition. But, even in this case, the difference between the medians was in the expected direction. The tendency to see the playing card in the Alley condition as more distant and larger than the card in the Reduced condition reflects the influence of the equidistance tendency upon the perceived distance, and thus the perceived size, of the playing card.

It should also be noted, however, that there was a tendency for the playing card to appear veridical in the absence of the visual alley. The calibrated distance score (5.5 ft.) of the card in the first presentation of the Reduced condition and its perceived size (2.5 in.) were similar to the simulated distance and width of 5 ft. and 2.25 in. respectively. The calibrated scores also indicate that although the playing card appeared to be almost twice as distant in the Alley condition as in the Reduced condition, it did not on the average appear at the back wall of the alley. If the card had appeared at the back wall, the average calibrated score from the Alley condition would have been 16.3 ft. But both the 10.3- and 12.2-ft. average results from the calibrated scores (Table I) are significantly different from 16.3 ft. at the .01 level ($t_{19} = 4.89$ and 6.29 respectively). Thus, as expected from the equidistance tendency, the playing card did appear *toward* the back of the alley but, consistent with the view that the final value is the resultant of all

the equidistance effects, did not, on the average, appear *at* the back wall of the alley.

The results from the estimates of the apparent distance of objects placed on the alley floor in the calibration situation are given in Table II, which shows that reported distance was a linear function of physical distance.

Discussion

Using the cue of optical contact, McDonald and O'Hara found a direct relation between the perceived size and perceived distance of triangles of a constant angular size.[10] A similar relation is present for the median data of Table I. The equidistance tendency resulting from the alley increased both perceived distance and perceived size of the experimental object. The fact that this occurred when a playing card was used indicates that the equidistance tendency has an effect despite the presence of familiar size as a cue to perceived size and perceived distance. Since the equidistance tendency normally is effective only when distance cues are reduced or ineffective, the present experiment suggests that familiar size is not a strong cue to either perceived size or perceived distance. On the other hand, the decrease in perceived size and distance of the card—a decrease which tended to occur between the first presentation of the Alley and the second presentation of the Reduced condition—suggests that in the absence of any opposing tendencies, familiar size can have a measurable effect.[11]

The view that the final equidistance effect on the playing card was the resultant of all the equidistance tendencies between the

TABLE II

Physical Distance (*D'*) and Reported Distance (*D'*): Calibration Situation

Object	*D* in ft.	*D'* in ft. Mean	Median	σ
Flashlight	3.0	2.2	2.0	1.4
Cigarette package	6.0	4.3	4.0	2.0
Glasses	8.0	5.8	5.0	2.8
Soup can	12.0	9.1	8.0	4.6
Butter carton	16.3	12.5	11.0	6.4

[10] McDonald and O'Hara, *op. cit.*, 276–280.

[11] A more recent experiment, however, indicates that a similar result can occur with rectangles of the same shape as playing cards, but without the familiar configuration. This result is attributed to a tendency to perceive objects at a near distance, in the absence of cues to depth. See W. C. Gogel, The sensing of retinal size, *Vision Res.*, in press.

card and other objects in the visual field makes possible the notion that this final effect may be graded in its influence on perceived distance. Thus, the equidistance tendency may be important in a variety of circumstances. For example, it has been applied in the explanation of the moon illusion.[12] If a disc of light (a moon) rather than a playing card had been used in the present study, it is clear that the disc would have appeared somewhere above the more distant parts of the alley. A change in the size and directional characteristics of the objects in the alley, or of the parts of the alley, would be expected to affect the final result and hence the apparent distance (and apparent size) of the disc. It is reasonable, therefore, as has been found, that the moon illusion should vary as a function of the characteristics of the terrain extending from *O* toward the moon.[13]

Summary

The image of a playing card was presented in an otherwise dark visual field or apparently suspended above the floor of a visual alley. The perceived size and perceived distance of the card were judged by 40 *O*s. In agreement with the expected effect of the equidistance tendency, the playing card appeared to be more distant and larger in the visual alley than in the reduced visual field. The results suggest that familiar size was a valid, but not a very effective, cue to perceived size and perceived distance. It seems that the equidistance tendencies occurring between simultaneous presentations of experimental objects and comparison fields can distort the perceived size and the perceived distance of an experimental object presented in the comparison field. Optical contact along a visual surface as a cue to depth was discussed as a limiting case of the equidistance tendency.

[12] Gogel, *op. cit., Psychol. Bull.*, 62, 1964, 159.

[13] L. Kaufman, and I. Rock, The moon illusion, *Sci. Amer.*, 207, 1962, 120–130; I. Rock and L. Kaufman, The moon illusion: II, The moon's apparent size as a function of the presence or absence of terrain, *Science*, 136, 1962, 1023–1031.

VISUAL DELAY AS A FUNCTION OF LUMINANCE

By J. A. Wilson and S. M. Anstis,
University of Bristol, England

It has long been known that dimly lit visual stimuli are seen later than bright ones, with a delay of up to 100 msec. This delay is responsible for the well-known Pulfrich pendulum stereo effect, and the amount of delay has been calculated from the magnitude of that effect.[1] In such experiments, a pendulum bob is set swinging in the frontoparallel plane. Neutral density filters are placed over one or both eyes, to dim *S*'s vision of the bob by a different amount for each eye. One eye sees later than the other, and since the bob is moving, the delay gives rise to a spatial binocular disparity which *S* interprets as a displacement in depth. The bob appears to swing in an ellipse; clockwise, as seen from above, when the denser filter is over the left eye, and counterclockwise when the denser filter is over the right eye. The more dense the filter, the greater the depth effect. By setting a pointer under the apparent near or far point of the swing, *S* can indicate the size of the depth effect. From this, the relative delay can be calculated: the more dense the filter, the greater the delay.

Our experiments gave a more direct measure of the delay, without using the stereo effect. We presented two targets, both seen by the same eye (or by both eyes), and controlled the intensity of each target separately. The two targets moved sinusoidally along parallel paths in the frontal plane, with a fixation point between. When they moved together, in phase, the dimmer target appeared to lag. *S* then adjusted the phase of one target until the two targets again appeared to move together. Thus he put in a phase lead which served to null the apparent phase lag. The setting was recorded. Preliminary experiments showed that the effect was a true time delay, not a phase lag: doubling the frequency of oscillation (from 1 cps to 2 cps) doubled the angular phase lag, and thus left the time delay unchanged.

* Received for publication February 11, 1969.

[1] Carl Pulfrich, Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie, *Naturwissenschaften*, 10, 1922, 553–564, 569–574, 596–601, 714–722, 735–743, 751–761.

Since our method presented both stimuli to the same eye, it showed incidentally that the delay is localized to position on the retina and cannot be a property of the eye as a whole. The delay cannot be due, for example, to any so-called state of adaptation of the whole eye; our results showed that the same retina responds with different delays to different illumination of its parts.

Method

Apparatus. Preliminary work showed that, as would be expected from Pulfrich-type experiments, the effect occurred both with a bright target on a dark ground and with a dark target on a bright ground. Using bright-on-dark, the effect could conveniently be shown on a suitable C. R. oscilloscope. In our main experiments, however, we used dark-on-bright, with marker shadows moving along back-illuminated slits. This allowed greater control over the conditions and a greater range of luminance. The results were essentially the same for both types of display, and only the latter are reported here.

The apparatus is shown in Fig. 1. *S* viewed binocularly through a reduction tube, a black card 45 cm. from his eyes. In the card were cut two parallel horizontal slits, 50 mm. long × 2 mm. wide, and spaced 20 mm. apart (approximately 6°, ¼°, 2½° visual angle). A fixation point was provided midway beween the slits. The slits and fixation point were backed by a sheet of ground glass, and back-illuminated by a projector.

Along each slit a spot shadow oscillated through a distance of 25 mm. (3°) at 1 cps. Each spot was the shadow cast by the projector from a vertical rod of 2-mm. diameter mounted as a crank on a synchronous motor. (To sharpen the shadows, the projector was slightly stopped down with a rectangular aperture in front of the lens). One rod was used for each shadow, each rod mounted on a separate motor. The two motors were mounted on the same vertical axis, facing one another. One had a clockwise rotation and the other counterclockwise, so that in this mounting they rotated in the same direction and in synchrony. The two shadows cast by the cranks then moved together, sinusoidally and in phase. To vary the phase of the upper spot, *S* could rotate the body of the upper motor, which was mounted in a bearing. The shaft of the motor rotated synchronously with respect to the body of the motor; hence a rotation of the body through a given angle rotated the shaft by the same angle additional to its synchronous rotation. The angular setting, of lag or lead, could be read from a protractor fixed to the body of the motor. The motors were checked frequently during the experiment to make sure that they had not jumped out of phase with each other as synchronous motors are apt to do.

Neutral density filters could be interposed over either slit, to a maximum of 6 log units. The maximum luminance at the slit (without filters) was 3.25 log ftL. (1800 ftL.). The minimum luminance used was $\bar{2}.25$ log ftL. (.018 ftL.).

Procedure. The apparatus was set up in a dark room, and *S* was dark-adapted for 30 min. before testing began. A filter was placed in front of the upper slit, to preset its luminance at one of five values: 3.25, 2.25, 1.25, 0.25, or $\bar{1}.25$ log ftL. For each of these preset values the lower slit was further reduced in luminance (relative to the upper slit) by 0, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, or

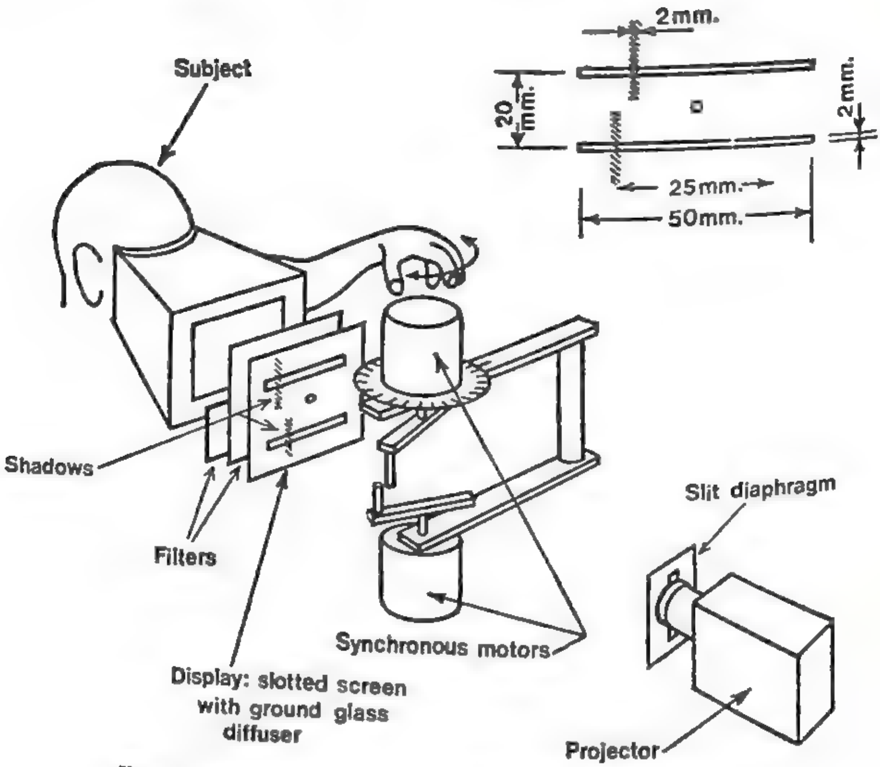


FIG. 1. APPARATUS AND, *Above Right*, DISPLAY SEEN BY *S*

4.0 log units. All of these conditions were used if *S* could see them; about 25 conditions in all were bright enough for the *S*s to see them and make settings. The conditions were presented in a systematic order from dimmest to brightest, since it seemed more important to avoid difficulties with adaptation than to avoid possible artifacts of order. Sometimes during the testing, *S* could observe afterimages; when this occurred, he was allowed time to dark-adapt. Thus *S* was always well adapted to the stimulus luminance presented.

RESULTS

The results for one *S* are shown in Fig. 2a (upper left corner). The curves show the delay of the lower marker shadow with respect to the upper, as a function of the luminance difference in log units. Each point is the mean of 10 readings. Similar results were obtained from the other two *S*s. For each curve of Fig. 2a, the upper slit had constant luminance, and the luminance of the lower slit was reduced from that value in steps. Thus the extreme right-hand point of each curve represents *S*'s setting when the luminance of the two slits was equal, and the other points represent *S*'s setting when the lower slit was dimmer and the upper slit the same as before; *i.e.*

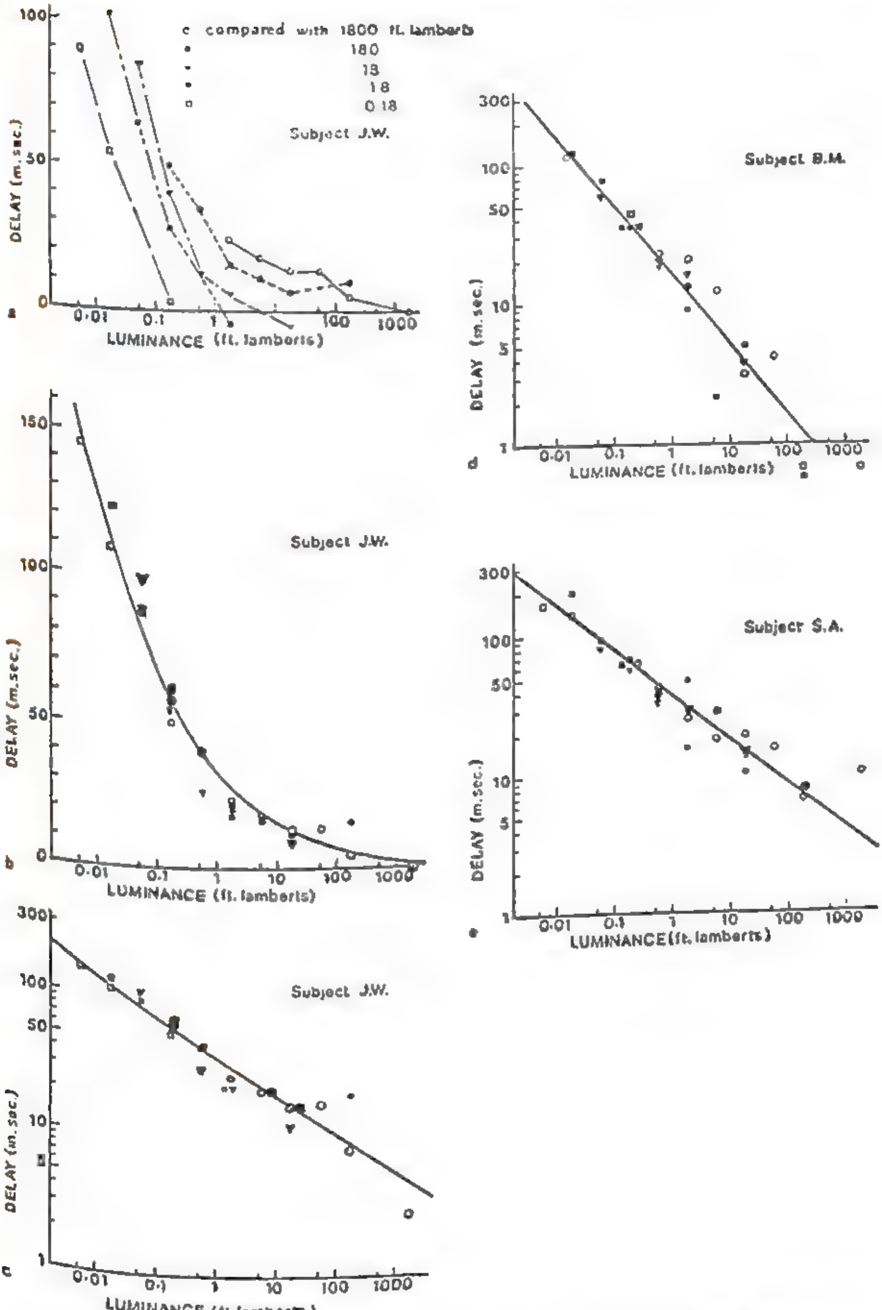


FIG. 2. VISUAL DELAY AS A FUNCTION OF LUMINANCE

(2a shows one S's delay for a dimly lit stimulus relative to a more intensely lit stimulus; 2b shows delay for stimuli of different luminance, referred to a luminance of 1800 ftL., points obtained from 2a; 2c shows data from 2b plus constant D_o plotted on logarithmic coordinates; 2d and 2e show logarithmic plots like 2c for the other two Ss.)

the standard luminosity to which the judgments on each curve were referred is given by the abscissa of the right-hand point of each curve.

The curves show that the movement on the dim slit appeared to lag behind that on the brighter slit and that the delay increased with increasing difference in luminance. Moreover, the curves show a steeper slope in dim light: for a given difference in luminance, the delay was greater when the overall luminance was low.

Treatment of results. Each curve of Fig. 2a shows the delay between one marker of fixed luminance and a second, dimmer marker of variable luminance. Each curve, however, covers only a small range of luminance. We could not extend the range, because if we did so, *S* was unable to see the dimmer marker: it was masked by the brighter. *S*'s eye could deal with 3 log units when light-adapted, and 1½ log units when dark-adapted.

However, it is possible to reconstruct from the data a single curve covering the whole of the range of luminance used in the experiments, by plotting *cumulative* delays. If a delay for 1 ftL. has been measured against a reference of 10 ftL., and a delay for 10 against a reference of 100, then we can add the two delays to get the delay for 1 against a reference of 100. If for each luminance there is a definite delay, independent of the luminance with which it is compared, then the cumulative delays between two luminosity levels would be the same, by whatever route they are calculated, and the set of cumulative delays would all lie on one curve. A least-squares test, described below, shows that our data points did lie on a smooth curve, so justifying the procedure for the conditions we used, *i.e.* with the eye adapted to each luminance which it viewed.

The cumulative-delay curve of Fig. 2b (center left) was obtained graphically by shifting the raw-data curves of Fig. 2a. Starting from the right (high luminosity, low delay), each curve in turn was slid upward by adding to it the delay appropriate to its right-hand (reference) luminance. This delay was obtained as the mean delay for the luminance on the curves to its right, already plotted. We had to suppose that even our brightest light had some delay (say D_o). To estimate the total cumulative delay we added D_o to the points on the curve, shifting it up even further. D_o was estimated by computing for it that value which gave the best-fitting curve to the data points. The curve fitted was a power function of the form:

$$D = aI^{-b} - D_o, \qquad [1]$$

where I is the luminance in ftL., D is observed delay in msec., D_o is an unknown small constant delay to be determined in fitting, and a and b are fitting parameters. The power-function term aI^{-b}, varying inversely with luminance, gives the major part of the observed delay. The residual constant term D_o allows the best-fitting power-law curve to be found and represents in principle the difference in delay between our brightest marker and a hypothetical infinitely bright marker, plus or minus our experimental error. (There may be other constant delays in perception, not affected by luminosity, but our experiment was inherently incapable of revealing these.)

The data and the fitted curve are shown in Fig. 2b (replotted from Fig. 2a) for one S. The fit was good; after allowing for the effects of shifting, the variance ratio was 13.3, $p < .001$. The equations fitted to the data for the three Ss were:

J.W.	$D = 36.00I^{-.29} - 2.7$ msec.	$V.R. = 13.3, p < .001$.
B.M.	$D = 16.71I^{-.50} + 1.8$ msec.	$V.R. = 31.0, p < .001$.
S.A.	$D = 37.20I^{-.32} - 8.0$ msec.	$V.R. = 4.5, p < .001$.

The power law held over the 5-log-unit range tested, from a barely visible light to an uncomfortably bright one. Over this range, for each tenfold reduction in intensity the delay increased by a factor of approximately 2 (J.W., S.A.) and 3 (B.M.). In other words, the delay varied roughly with the inverse square or cube root of the luminosity. The longest delays were of the order of 100 msec. (observed) and 150 msec. (cumulative).

A power-law curve when plotted on logarithmic coordinates becomes a straight line, with slope corresponding to the exponent. The data of Fig. 2b are shown replotted in Fig. 2c with log coordinates on both axes. Before plotting, D_o was added to each of the data points; this shifts the origin to take account of residual delay, leaving only the power-law term in the fitted curve. This, when plotted on logarithmic coordinates, becomes a straight line. The data for the other two S are shown in Fig. 2d and 2e, treated in the same way.

Discussion

Our data showed that visual perception is subject to delay, dimly lit objects being seen later than those which are strongly lit. When all the delays were referred to the same standard luminosity, the data fit a power-law curve, $D = aI^{-b} + D_o$. The measurements were of delay in one percept relative to another of different luminance, but the fitted curve represents cumulative delay, relative to the greatest luminance which we used. The greatest measured delay was

100 msec., and the delays accumulated to 150 msec. These values are much greater than those reported by workers using Pulfrich-type experiments to measure delay, although much less than the retinal latencies reported by researchers in animal physiology.

The delays as measured from our three *S*s were of very nearly the same magnitude, but the exponent b of the fitted curve, controlling the slope of the logarithmic plot, was rather different in one of the *S*s: −0.5 for B.M., as against −0.29 for J. W. and −0.32 for S.A. This difference was probably due to experimental error, the fitted value of b being rather sensitive to small errors at the high-luminance, short-delay end.

Comparison with data of Lit[2] and others. Lit[3] and Rock and Fox[4] have reported sensory experiments on the Pulfrich stereo phenomenon, relating luminance to perceptual delay; and Adrian and Matthews[5] and Hartline[6] have reported physiological experiments on retinal and optic-nerve latencies. With the exception of Rock and Fox, the data given by all of these writers can be fitted to a power function.

Lit measured the effect of binocular differences of luminous intensity on the magnitude of the Pulfrich stereo effect and from his data calculated the effect of intensity on delay. Lit's experiments, like ours, covered a fairly wide range of luminance in a number of overlapping segments, and his results were plotted as a number of short curves, each representing measured delay. He suggested that his results would be explained if delay were an inverse function of log intensity, but he did not calculate cumulative delays as we have done. Replotted in this way (Fig. 3), Lit's data fit a power function remarkably well. The best fitting equations are:

$$\text{C.G.M.} \quad D = 71.3I^{-.14} + 15.5 \text{ msec.}$$
$$\text{A.H.} \quad D = 74.0I^{-.19} + 13.0 \text{ msec.}$$

The only marked discrepancy from our findings is that the slope of the plotted line is much less steep for Lit's subjects (−0.14 and −0.19) than for ours (−0.29, −0.32, −0.5). We cannot yet ex-

[2] Alfred Lit, The magnitude of the Pulfrich stereophenomenon as a function of binocular differences of intensity at various levels of illumination, this JOURNAL, 62, 1949, 159–181.

[3] Lit, *op. cit.*, 171.

[4] M. L. Rock and B. H. Fox, Two aspects of the Pulfrich phenomenon, this JOURNAL, 62, 1949, 279–284.

[5] E. D. Adrian and Rachel Matthews, The action of light on the eye, *J. Physiol.*, 63, 1927, 378–414; 64, 1927–28, 279–301.

[6] H. K. Hartline, The discharge of nerve impulses from the single visual sense cell, *Cold Spring Harbor Sympos. quant. Biol.*, 3, 1935, 245–250.

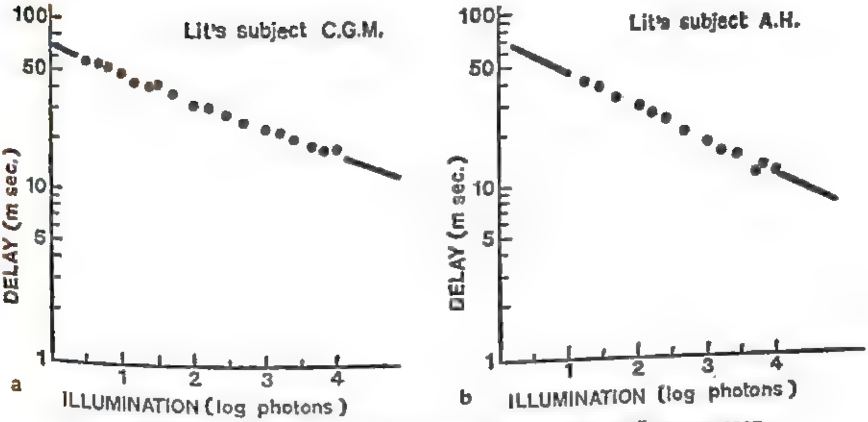


FIG. 3. VISUAL DELAY FOR STIMULI OF DIFFERENT LUMINANCE, REPLOTTED FROM LIT'S DATA

plain this difference in slope. Lit used artificial pupils, and we did not; but the effect of variation in pupil size would be to reduce the slope, making our results more like Lit's.

Rock and Fox's conclusions are inconsistent with ours. They give data to which they fit the function: stereo effect = k log log (1/transmission). With the usual method of conversion, delay is approximately proportional to stereo effect over the range which they covered, and the equation is then equivalent to $D = k \log \log (1/I)$. Neither our data nor Lit's could fit such a law. However, the data given by Rock and Fox cover a range of less than 2 log units of luminance, whereas Lit's data cover 3.5 log units and ours cover 5.5. It seems possible that had they extended the luminance range of their experiment, their data would no longer have fitted the log-log law. Also, the delays calculated by the usual method from Rock and Fox's data are very small: at 2.7 ftL. their delay was 4 msec. (vs. our 10 msec.) and at 0.06 ftL. their delay was 10 msec. (vs. our 70 msec.). None of their delays was longer than 10 msec. (vs. our maximum delay of 100 msec.) This discrepancy leads us to suppose that delay due to luminance differences is not the only important influence on the stereo effect which they measured and that our application of the usual conversion to their results may not be justified.

It is worth noting that physiological preparations show retinal latencies which vary inversely with illumination and that a power function can be fitted to such data. The data given by Adrian and Matthews[7] for retinal delay in the eel fit a power function:

[7] Adrian and Matthews, *op. cit.*, 292, Fig. 8.

$$D = a_1 I^{-.22} - 5.9 \text{ msec.} \quad p < .01 .$$

The constant a_1 is undetermined since Adrian and Matthews used arbitrary light units. This equation results in a better fit than the reciprocal curve which they give as a first approximation. A figure published by Hartline[8] for optic-nerve delay in the frog yields data which also fit a power function:

$$D = a_2 I^{-.75} + 250 \text{ msec.} \quad p < .01 .$$

Again the constant, a_2, is undetermined. The 250 msec. obtained for D_o is large, but Hartline was measuring optic-nerve delay, not retinal delay. Adrian and Matthews had allowed 60–90 msec. for retinal-to-nerve delay, and if this were replaced in the data the two equations would be more nearly comparable. However, the comparison should not be pressed too far. The data come from different species, and the retinal latencies measured physiologically were very much greater than the subjective delays which were measured psychophysically in the present study.

[8] Hartline, *loc. cit.*

LIGHTNESS AND ORIENTATION

By Jacob Beck, University of Oregon

Studies of lightness perception have shown that perceived lightness may be affected by the apparent position of a surface.[1] It has been further suggested that a variation in luminance may be seen as a shadow (a difference in the illumination of a surface) or as a gray spot (a difference in the lightness of a surface) consistent in each case with the apparent position of the surface.[2] The present paper was based on an observation which provides further information about the way in which the orientation of a surface may affect perceived lightness. The informal observation was that when one looked at identical photographs of a young girl in her bathing suit, the shadowed chest area between the girl's chin and halter appeared darker in the inverted photograph than in the upright photograph (Fig. 1; note, however, that because of reduction and distortions occurring in printing, this effect may not be clear in the reproduced photographs). Six experiments were undertaken to examine the conditions which affect this judgment.

Method

The general procedure was the same in each experiment. Two carefully matched photographs, one upright and one inverted, were mounted three inches apart on black cardboard. The *O*s were asked to judge whether corresponding areas in the photographs were equal in lightness, or whether one area was lighter or darker than the other. The *O*s were told that the photographs were made as equal as possible and that the purpose of the experiment was to test the accuracy of the photographic reproductions. The photographs were uniformly illuminated by a fluorescent lamp on the ceiling behind *O*. The illumination of the photographs was 31 ftc. To insure that no bias was introduced by a particular photograph, the targets were inverted for alternate *O*s,

* Received for publication December 30, 1968. The research reported was supported by National Science Foundation Grant GB-5285 to the University of Oregon.

[1] R. S. Woodworth and H. Schlosberg, *Experimental Psychology*, 1959, 444; G. Katona, Color contrast and color constancy, *J. exp. Psychol.*, 18, 1935, 49–63; E. Mach, *The Analysis of Sensations*, 1959, 209; J. E. Hochberg and J. Beck, Apparent spatial arrangement and perceived brightness, *J. exp. Psychol.*, 47, 1954, 263–266; J. Beck, Apparent spatial location and the perception of lightness, *J. exp. Psychol.*, 69, 1965, 170–179.

[2] Beck, *loc. cit.*

FIG. 1. A PHOTOGRAPH FROM R. M. EVANS, *Eye, Film, and Camera,* 1959, P. 279, AND AN IDENTICAL INVERTED

so that the photograph upright for one O became the inverted photograph for the next O. Both photographs were 6.5×8 cm. New Os served in each experiment.

In *Experiment I,* the Os were seated 90 cm. from the photographs, which were set upright on a table. They were asked to judge in turn whether three corresponding areas on the two photographs were equal, darker, or lighter in lightness. The areas were (*a*) the white rim of the bucket behind and to the left of the child, (*b*) the shadowed chest area between the chin and halter, and (*c*) the black cloth on the ground to the left of the child. The three areas were pointed to by E in a different irregular order for each O. Following their judgments of lightness, the Os were asked whether the space between the child's chin and chest appeared equal in the two photographs or whether the space appeared to be greater or less in one of the pictures; *i.e.* they were asked to make judgments of depth. Twenty Os served in the experiment.

In *Experiment II,* the arrangement and procedure were identical with those of Experiment I. The Os, however, now viewed the pictures from a distance of 210 cm. Twenty Os served in the experiment.

In *Experiment III,* the pictures were cropped from just below the girl's nose (the upper lip was visible) to below her halter and included the whole of the shadow under the chin. On the left and right, the photographs just included the straps of her halter. The Os viewed the upright and inverted photographs were mounted as in Experiment I. The Os were asked to compare two areas of the photographs: (*a*) the shadowed chest area between the chin and the halter, and (*b*) the white area of the girl's left shoulder between her left cheek and the halter strap. The order in which the Os made judgments of lightness of the two areas was alternated. Following these judgments the Os were asked to compare the perceived space between the child's chin and chest in the two photographs. The remaining

procedure and arrangement were identical with those of Experiment I. Twenty *O*s served in the experiment.

In *Experiment IV*, the photographs were cropped so that *only* the shadowed area between the chin and halter was visible. The two targets were again mounted upright and inverted. The *O*s viewed the targets from a distance of 90 cm. The *O*s were asked to judge the lightness of the shadowed areas. Twenty *O*s served in the experiment.

In *Experiment V*, the same photographs were used as in Experiment I. The observation distance was 90 cm. The *O*s were asked to judge the lightness of the shadowed area between the chin and halter. They were also asked to compare the perceived space between the child's chin and chest in the two photographs. Each *O*, however, viewed the photographs from two different positions: (*a*) with his head upright, as in Experiment I, and (*b*) with his head inverted, *O* looking between his feet at the photographs. The order of the two positions from which *O*s viewed the photographs was alternated. Twenty-five *O*s served in the experiment.

In *Experiment VI*, the cropped photographs used in Experiment IV (those showing only the shadows) were again the targets. The *O*s viewed the targets with their heads inverted, as in Experiment V. They were asked to judge the lightness of the two shadowed areas. Twenty *O*s served in the experiment.

Results

Table I presents the results of the six experiments. The numbers of lighter, darker, and same judgments of the corresponding areas in the two photographs are shown in terms of the upright photograph. Except for one condition of Experiment V and in Experiment VI, when the *O*s viewed the photographs with heads inverted, the upright photograph was also upright with respect to an *O*'s eyes; in the two excepted instances, the upright photograph was inverted with respect to an *O*'s eyes and the inverted photograph was upright with respect to an *O*'s eyes.

The results indicate that in Experiments I, II and III, the chest area was judged lighter a significantly greater number of times when the photograph was presented upright than when it was presented inverted. For example, in Experiment I, 16 of the 20 *O*s saw the chest area of the upright photograph as lighter, one *O* saw it as darker, and three *O*s saw the lightness of the two areas as the same. Increasing the distance of observation in Experiment II did not markedly reduce the tendency to see the chest area as lighter when the photograph was upright. Experiment IV, however, showed that when all information was eliminated and the gray area of the shadow alone remained, there was no tendency to see the shadowed area as lighter or darker depending on the photograph's orientation. In Experiment V, the results confirmed those

TABLE I

LIGHTNESS JUDGMENTS

	Upright lighter	Upright darker	Same		Upright lighter	Upright darker	Same
Experiment I				Experiment II			
Chest	16*	1	3		12*	1	7
Cloth	5	2	13		4	2	14
White rim	3	7	10		5	7	8
Experiment III†				Experiment IV			
Chest	13*	2	5		7	6	7
Shoulder	2	7	9				
Experiment V				Experiment VI			
(Head upright) Chest	18*	1	6				
(Head inverted) Chest	3	11	11		4	8	8

* $p < .01$ (two-tailed test).
† In Experiment III, two *Os* could not see the shoulder clearly enough to judge its lightness.

of Experiment I when the *O*s viewed the photographs with their heads upright. When the *O*s viewed the photographs with their heads inverted, however, the inverted picture, which was thus upright with respect to their eyes, tended to be seen as lighter. A sign test of the difference in the number of *O*s who reported the chest area of the upright photograph as lighter when their heads were upright (18) compared to when their heads were inverted (3) was significant at the .01 level. Thus, the orientation of the shadow gradient with respect to the eye was a more important factor in determining the perception of lightness than was the orientation of the shadow gradient with respect to the actually upright or inverted position of the photograph. Experiment VI confirmed the results of Experiment IV: when only the shadowed area was presented, there was no systematic tendency to see either the upright or inverted photograph as lighter.

Table II presents the *O*s' judgments of depth. In Experiments I, II and III, greater depth was seen when the photograph was upright than when it was inverted. For example, in Experiment I, 17 of the 20 *O*s reported that the inverted photograph appeared to be flatter, and three *O*s reported that the space between the chin and the chest in the two appeared to be the same. In Experiment V, the results were similar when an *O*'s head was upright. When an *O*'s head was upside down, however, greater depth was seen in the photograph that was inverted and thus oriented normally with respect to *O*'s head and eyes. Table II shows that sign tests of the differences in the numbers of *O*s who reported greater depth when the photograph was upright compared to when the photograph was inverted with respect to *O*'s eyes were significant at the .01 level in all experiments.

Discussion

The results indicate that the child's chest was perceived to be a lighter gray when the photograph was upright with respect to

TABLE II

Depth Judgments

	Upright deeper	Upright flatter	Same
Experiment I	17*	0	3
Experiment II	15*	2	3
Experiment III	14*	1	5
Experiment V			
Head upright	15*	2	8
Head inverted	2*	13	10

* $p < .01$ (two-tailed test).

O's eyes than when the photograph was inverted. Also, greater depth was perceived between the child's chin and chest when the photograph was upright with respect to O's eyes. Since the same luminance relationships evoked different perceptions of lightness and orientation, the question raised by these results concerns the way in which the perception of relative lightness and orientation in the two photographs was modified by central factors.

Careful observation of the photographs suggests an important difference between the upright and inverted positions. In the upright position the photograph of the child looks normal. When the photograph is turned upside down, there is a disorganization of the percept. For example, the pattern of light and shade on the face is no longer seen as belonging to one surface but tends to be seen as separate areas unrelated to each other. Many Os also reported that the dark area at the base of the child's neck appeared much clearer when the photograph was inverted with respect to their eyes than when it was upright. In the upright position this black band was indistinct and O had to look for it. When the photograph was inverted, the black band was prominent and emerged at once.

An earlier study suggested that when a target reflects a distribution of luminances, the perceived lightness of a surface depends upon a prior decision as to whether a surface is of a single lightness or not: if the cue properties of stimuli lead to the decision that the surface is of a uniform lightness, a single lightness is fitted to a varying luminance pattern with deviations perceived as differences in orientation and illumination of the surface; if the decision is that the surface consists of multiple lightnesses, the varying luminance pattern then determines differing lightnesses.[3] The same paper further indicated that the lightness fitted to a luminance pattern is not unique and can depend on the apparent orientation of a surface. Though limiting the subsequent integrations that may occur, a particular luminance pattern can be indeterminate, evoking a perception of a lighter surface with shadows when the surface is seen turned away from the light or a darker surface with light spots when the surface is seen as facing the light. What is seen as a shadow, a light spot, and a normally illuminated surface is made consistent in each case with the apparent position of the surface.

A similar explanation may be applied to the present results.

[3] Beck, *loc. cit.*

When the photograph is upright one tends to see the chest and face of the child as single surfaces of a specific lightness. Variations of luminance are then seen as shadows and highlights. For example, one sees areas of the child's face which are illuminated and areas which are shadowed due to differences in the orientation of these areas to the light source. When the photograph is inverted with respect to one's eyes, the relatedness of the areas changes. The photograph is seen as made up of separate areas and the image appears more flat; as separate areas, the variations of luminance are seen as differences in lightness. The figural organization of the pattern may also exert a selective effect which emphasizes some part or aspect of the percept; *e.g.* the band at the base of the neck is clearer when the photograph is seen as a surface made up of multiple lightnesses than when it is seen as a surface with a single lightness. Thus, the difference between judgments of the upright and inverted photographs is not a simple response to the information in the photographs. Both are clearly pictures of a child, and *O*'s most likely hypothesis about the darkened area is that it is a shadow due to the orientation of the child's body with respect to a light source. Perception, however, depends on highly specific information such as the orientation of the stimulus with respect to the eye.[4] When the photograph is upright, the lightness variations apparently evoke a schema which leads *O* to see the parts of the photograph as surfaces of a single lightness varying in orientation and illumination. When the photograph is inverted, this schema is not evoked and the areas are seen as separate and differing in reflectance.

It is also important to consider the fact that the difference in orientation only slightly modifies the lightness quality. The apparent position of a surface often produces clear differences in perceived lightness.[5] Photography, however, is not a favorable medium for representing differences in illumination. Because of contradictory stimulation indicating that the image is flat and because of the presence of white borders on a photograph, there is a strong tendency for every variation in luminance to evoke the perception of gray.[6] Shadows and highlights in a photograph are, therefore, ordinarily seen as differences in gray rather than as differences in

[4] The effect may also depend on inhomogeneity. A print in which the inhomogeneity in the shadowed chest area was reduced gave more variable results.

[5] See n. 1.

[6] R. M. Evans, *Eye, Film, and Camera*, 1959, 162–167.

illumination. Thus, differences in organization which would ordinarily give large lightness differences give only slight differences in photographs. The importance of the present demonstration is that the changes in lightness, though slight, provide further information about how central processes determine perception of lightness.

Summary

When a photograph was upright with respect to an *O*'s eyes, an area of reduced luminance on the photograph was perceived to be lighter than when the photograph was inverted. To explain these results, it was proposed that variations of luminance are analyzed differently depending upon a prior decision as to whether the surface is of a single lightness or not. If the cue properties of stimuli lead to the decision that the surface is of a uniform lightness, a single lightness is fitted to a varying luminance pattern with deviations perceived as differences in orientation and illumination of the surface. If the decision is that the surface consists of multiple lightnesses, the varying luminance pattern then determines differing lightnesses.

HOW IMPLICATION IS UNDERSTOOD

By P. N. JOHNSON-LAIRD and JOANNA TAGART,
University College London, England

Conditional sentences of the form "if p then q" are often difficult to evaluate, as students of logic well know. The conditional is clearly true when both antecedent (p) and consequent (q) are true, and false when the antecedent is true but the consequent false. But what truth-value should be assigned when the antecedent is false? Logicians, working with a propositional calculus that permits only values of truth or falsity, stipulate that the implication is true in this case—regardless of the truth-value of the consequent. However, it seems that the conditional sentence might fail to do justice to this notion of *material* implication.

Conditional sentences have been found to present difficulties to both children and adults. Matalon and Peel suggest that children tend to interpret the conditional as a material equivalence ("if and only if p then q"), which is true when antecedent and consequent have the same truth-value and false when their values are different.[1] Wason, however, argues that adults do not treat the conditional in a truth-functional manner: they consider it to be *irrelevant* when its antecedent is false.[2] For example, when someone says, "if it's raining, then I'm going to the cinema," the statement is neither true nor false but merely irrelevant if in fact it is *not* raining. The layman is unlikely to consider the statement to be true just because it is not raining. Indeed, logicians have long recognized, and argued about, this way of interpreting conditional sentences.[3]

The question naturally arises as to the extent to which the inter-

* Received for publication August 20, 1968. The authors express their gratitude to Dr. P. C. Wason for his advice and encouragement, and for a critical reading of an earlier version of this paper.

[1] B. Matalon, Etude genetique de l'implication, *Etudes d'epistemologie genetique: XVI, Implication, formalisation et logique naturelle*, 1962, 69–95; E. A. Peel, A method for investigating children's understanding of certain logical connectives used in binary propositional thinking, *Brit. J. math. statist. Psychol.*, 20, 1967, 81–92.

[2] P. C. Wason, Reasoning, in B. M. Foss (ed.), *New Horizons in Psychology*, 1966, 135–151.

[3] W. V. O. Quine, *Methods of Logic*, rev. ed., 1956, 12; W. Kneale and M. Kneale, *The Development of Logic*, 1962, 128–138.

pretation of implication is influenced by the manner of its expression. The present experiment was designed to examine the effect of statements of these types:

1. If p then q.
2. There isn't p, if there isn't q.
3. Either there isn't p, or there is q (or both).
4. There is never p without there being q.

To the logician, these sentences may be interpreted as expressing material implication.[4] On Wason's hypothesis, however, we expected that Sentence 1 would be considered irrelevant to any situation that falsified its antecedent. Similarly, Sentence 2, which was derived from the contrapositive "if not q, then not p," would also be considered irrelevant when its antecedent is false, *i.e.* when q is true. But Sentences 3 and 4 were not conditionals: they lack the conditional term *if*, and it was predicted that they would be treated as expressing material implication and elicit fewer judgments of irrelevancy. It will be noted that Sentence 3 contains the term *either* which tends to suggest exclusive disjunction[5] but that this is countermanded by the presence of *or both*, which is an explicit statement of inclusive disjunction. The predicted classifications are summarized in Table I.

There are a number of ways in which an antecedent or consequent may be falsified. For example, an antecedent like "if there's a letter A," which was used in the present experiment, can be falsified by the occurrence of a letter B, or of a geometrical shape, or of nothing whatsoever. Logically, these would be equivalent falsifications; psychologically, they might not be equivalent. This point was examined by using a suitable selection of stimuli.

TABLE I
PREDICTED CLASSIFICATIONS OF THE FOUR TYPES OF SENTENCE

Sentence type	Situation			
	pq	$p\bar{q}$	$\bar{p}q$	$\bar{p}\bar{q}$
1. If p then q.	T	F	?	?
2. Not–p if not–q.	?	F	?	T
3. Not–p or q.	T	F	T	T
4. Never p without q.	T	F	T	T

Note: T, F, and ?, respectively, denote judgments of truth, falsity, and irrelevance; $\bar{p}$ denotes not–p.

[4] P. F. Strawson, *Introduction to Logical Theory*, 1952, 35–40.
[5] A. Naess, L'emploi de la disjonction chez les adolescents, *Etudes d'epistemologie genetique: XVI, Implication, formalisation et logique naturelle*, 1962, 151–158.

Method

Each *S* was shown an array of stimuli and a sentence referring to them. The *S*'s task was to consider each stimulus in turn and to decide whether it indicated that the sentence was true or false, or was irrelevant to the truth-value of the sentence. *S*s served as their own controls and performed the task with the four different sentences expressing implication. The order of presentation of the sentences was counterbalanced over *S*s: each of the 4! possible orders was used once.

Materials. The sentences presented were variations of the four basic types: *e.g.* (1) If there is an A on the left, then there is a 7 on the right. (2) There isn't an A on the left, if there isn't a 7 on the right. (3) Either there isn't an A on the left, or there is a 7 on the right (or both). (4) There is never an A on the left without there being a 7 on the right. In order to reduce residual effects, the numbers and letters in each sentence that *S* received, and the corresponding arrays of stimuli, were different. There were four such sets of material, involving different letters and numbers, and their order of presentation was independently counterbalanced over the *S*s. The 16 sentences (4 types × 4 contents) were typed in capital letters on separate 6 × 2 in. cards.

Each set of stimuli consisted of 16 4 × 2 in. cards which were divided into two halves by a heavy ink line. On the left of the line there was a letter (either the one mentioned in the sentence or one other letter), or a geometrical shape, or nothing whatsoever; on the right of the line there was a number (either the one mentioned in the sentence or one other number), or the geometrical shape, or nothing whatsoever. Four such sets of stimuli were constructed with different numbers, letters, and shapes.

Procedure. The first set of stimuli was spread out in an arbitrary array in front of *S*, and the general purpose of the experiment was described. The *S* was told that although any stimulus must fall into one of the three classificatory categories ('true,' 'false,' 'irrelevant'), he must not assume that there would necessarily be cards in all three categories. When *S* understood what he had to do, the stimuli were gathered together and shuffled. *S* was told that he was going to classify the cards one at a time, each card being placed in an appropriate pile, and that he would be timed. *S* was timed from the moment that he received the sentence until he had completed the classification of the stimuli. The sentence remained on view throughout the classification. The subsequent classifications followed the same procedure except that the initial display of stimuli was omitted.

Subjects. Twenty-four *S*s were individually tested. They were all students at University College London, and native speakers of English.

Results

There were four stimuli which were crucial in evaluating the results. They consisted, for a sentence of the form "if there's an A . . ., then there's a 7 . . .," of the items A7, A8, B7, and B8, *i.e.* pq, $p\bar{q}$, $\bar{p}q$, and $\overline{pq}$. Table II shows the six common classifications of these stimuli, and the frequencies with which they occurred for each type of sentence. None of the remaining classifications occurred

more than twice throughout the whole experiment. There were 28 different types of classification altogether, 6 common and 22 miscellaneous ones, out of a total of 81 possible classifications.

The most frequent classification, for Sentences 1, 2, and 3, was the predicted one; and the actual frequencies were all significant ($p < .01$), assuming independent classifications, on binomial tests based conservatively upon the actual number of different types of classification for each sentence. Contrary to expectation, Sentence 4 tended to be classified in the same way as Sentence 1. Fifteen *S*s produced the same classification for these two sentences, whereas there were only four other occasions when an *S* produced the same classification for two sentences. The chance probability of obtaining the same classification for Sentences 1 and 4, with 2 and 3 being different and different from one another, is conservatively 1/28. Clearly, the similarity of the classifications of Sentences 1 and 4 was not due to chance.

The mean number of 'irrelevant' judgments of the 16 stimuli and the mean classification times are given in Table III. The difference between the sentences of the number of 'irrelevant' judgments was significant on a Friedman analysis of variance ($\chi^2_r = 32.3$, $p < .001$). Sentences 1 and 4 tended to elicit 'irrelevant' judgments when their antecedents were false, and this was also the case to a lesser extent for Sentence 2. There was no tendency in any condition for 'irrelevant' judgments to increase when falsification was due to a geometrical shape or a 'blank' rather than to a letter or a number. On this point the *S*s' rationality was vindicated.

The classification times for the four sentences were also significantly different on a Friedman analysis of variance ($\chi^2_r = 34.5$,

TABLE II

Classifications of the Four Crucial Stimuli and Classification Frequencies of Occurrence for Each Type of Sentence

	Stimuli				Sentence type				
Common classifications	pq	$p\bar{q}$	$\bar{p}q$	$\bar{p}\bar{q}$	1	2	3	4	Total
	T	F	?	?	19	1		14	34
	T	F	T	T	1	1	8	3	13
	?	F	?	T		5			5
	F	F	T	T		2	4		6
	F	F	T	?		1	2		3
	T	F	?	T		2		1	3
Miscellaneous classifications					4	12	10	6	32
Totals					24	24	24	24	96

Note: Miscellaneous classifications are those which did not occur more than twice throughout the whole experiment.

TABLE III

MEAN NUMBER OF 'IRRELEVANT' JUDGMENTS AND MEAN CLASSIFICATION TIMES FOR EACH TYPE OF SENTENCE

	Sentence type			
	1	2	3	4
'Irrelevant' judgments	10.2	5.1	1.6	9.3
Classification times (in sec.)	45	96	107	60

$p < .001$). Although there was a significant learning effect for classification times ($p < .003$, Jonckheere group test for predicted trend), there was no such effect for logical accuracy.

DISCUSSION

This experiment showed that the way in which implication is expressed exerts a decisive influence upon what it is understood to denote. When expressed in the form of a conditional "if p then q" or "not p if not q" it was, as predicted, treated in a non-truth-functional manner. Unexpectedly, the same interpretation—in which stimuli falsifying the antecedent were regarded as irrelevant—was elicted by the sentence "there is never p without q." Hence, the term *if* is by no means necessary to elicit the non-truth-functional interpretation; and in the absence of an account of the semantics of these sentences, such necessary conditions remain obscure. Similarily, it seems likely that *if* cannot be taken as an unequivocal marker of the antecedent of conditionals: a sentence of the form "p only if q" is likely to receive the same interpretation as "if p then q." To what extent is the non-truth-functional interpretation due to the implicit invitation to classify stimuli as irrelevant? Performance on the disjunctive sentence suggests that *S*s were able to resist the 'irrelevant' category, and it is plausible to assume that the classifications did reflect the spontaneous interpretations of the sentences. Likewise, it seems improbable that the specific content of the sentences and stimuli should have exerted any major distorting influence upon performance.

The sentence which was most often classified as material implication was the disjunction. This was never treated as exclusive disjunction, but an interesting error proved most persistent. Such was the force of the phrase "either there isn't p . . ." that a number of *S*s produced a truth-functional classification appropriate to the simple proposition "not p" (see Table II).

The manner in which the conditionals were interpreted, considered

in conjunction with the findings of Matalon and Peel,[6] raises certain difficulties for Piaget's account of intellectual development. Preadolescent children tend to treat conditionals as expressions of material equivalence; adolescents at the level of propositional operations treat them as expressions of material implication; yet, undergraduates in the present experiment failed to treat them as any sort of truth-functional connective. Piaget believes that an individual tests a putative causal relation by expressing it in the form "if p then q" and then searching for its counterexample, formed by negating the material implication.[7] But adults evidently do not readily interpret "if p then q" as material implication. Even if they did, further doubt is cast upon Piaget's position by an unpublished experiment by Wason and Johnson-Laird, in which subjects were presented with four cards bearing values of p, $\bar{p}$, q, and $\bar{q}$. They were told that every card had a value of p or $\bar{p}$ on one side, and q or $\bar{q}$ on the other side; and they were asked to choose those cards which it was necessary to turn over to test whether a given conditional rule was true or false. There was, indeed, a consistent tendency for subjects to choose the cards which fulfilled the antecedent: p in the case of "if p then q," and $\bar{q}$ in the case of "if not q then not p." However, subjects were reluctant to choose those cards which falsified the consequent, especially in the case of "if p then q." Yet such cards are a required choice on any reasonable interpretation of the conditional, including even the non-truth-functional interpretation of the present experiment.

Such a result makes a stark contrast with Piaget's views, and with the findings of Stewart and Hill that adults and children correctly evaluate inferences of the form "if p then q; not q, therefore not p."[8] It would seem therefore that there are crucial psychological differences between *making* inferences and merely *evaluating* them. Not only do *Ss* fail to make the inference in the card-turning test, but their failure, as Wason has shown, is resistant to a number of "therapeutic" procedures.[9]

Finally, we may ask how implication is best expressed in the

[6] Matalon, *loc. cit.*; Peel, *loc. cit.*

[7] E. W. Beth and J. Piaget, *Mathematical Epistemology and Psychology*, 1966, 181.

[8] D. K. Stewart, Communication and logic: Evidence for the existence of validity patterns, *J. gen. Psychol.*, 64, 1961, 297–305; S. Hill's findings reported by P. Suppes, On the behavioral foundations of mathematical concepts, in L. N. Morrisett and J. Vinsonhaler (eds.), Mathematical learning, *Monogr. Soc. Res. Child Dev.*, 30, 1965 (No. 1), 60–96.

[9] P. C. Wason, Reasoning about a rule, *Quart. J. exp. Psychol.*, 20, 1968, 273–281.

English language. There is no readily available answer to this question: we are faced with a dilemma. On the one hand, disjunction yields an implicational interpretation more often than the conditional sentences, but it takes longer to process and has a tendency to produce diverse and labile interpretations—a finding which has been recently confirmed.[10] And such ambiguities are likely to be reflected in tasks involving the evaluation of inferences.[11] On the other hand, performance with "if p then q," though faster and more stable, is not consistent with material implication. However, this departure from the logicians' calculus has an unexpected advantage. It breaks the logical relation between the conditional and its contrapositive: they no longer imply one another. This does away with the paradoxes of material implication[12] and the paradoxes of confirmation,[13] at least for the conditionals of everyday language.

Summary

Students classified stimuli according to whether they indicated that a sentence was true or false, or were irrelevant to the truth-value of the sentence. Four different sentences were used, with Ss acting as their own controls, and each sentence was logically equivalent to a material implication. The results showed that disjunction ("not-p or q") yielded the greatest number of classifications in accordance with the truth-values of implication. The remaining sentences ("if p then q," "not-p if not-q," "never p without q") were not classified in a truth-functional way: stimuli were judged irrelevant when they falsified the antecedents of these sentences. The results would seem to raise some difficulties for Piaget's notion of the developmental level of formal operations.

[10] P. C. Wason and P. N. Johnson-Laird, Proving a disjunctive rule, *Quart. J. exp. psychol.*, 21, 1969, 14–20.

[11] P. N. Johnson-Laird, On understanding logically complex sentences, *Quart. J. exp. Psychol.*, 21, 1969, 1–13; P. N. Johnson-Laird, Reasoning with ambiguous sentences, *Brit. J. Psychol.*, 60, 1969, 17–23.

[12] Strawson, *op. cit.*, 88.

[13] C. I. Hempel, Studies in the logic of confirmation, *Mind*, 54, 1945, 1–26, 97–121, reprinted in C. I. Hempel, *Aspects of Scientific Explanation*, 1965, 3–46.

FORECASTING EVENT CYCLES: A LABORATORY PARADIGM

By Thornton B. Roby, Tufts University

The recognition of recurrent sequences of natural events is an important part of man's adjustment to his environment. In simple economies, hunting, agriculture, and husbandry are meshed directly with seasonal and biological cycles. As societies become more complex, the behavior of each individual must be synchronized, in some degree, with regularities in the activity of his associates,[1] so that there are many distinct cycles to keep track of and respond to. Moreover, the recurrence of event sequences is rarely exact. Sunrise and sunset occur at different times depending on the season, biological cycles are modified by exogenous factors such as weather, and the secondary cycles of social activity are even more subject to change. Knowledge of differences among these variants may be as vital for successful adjustment as recognition of the cycles themselves.

The difficult conceptual problem of recognizing and memorizing natural cycles may, in principle, be mitigated in several ways: *(1) Compounding.* If a cycle consists of complex events, the elements of that complex may undergo separate identifiable cycles which are simpler than the overall cyclic pattern. The prediction of tides or eclipses provides a clear example. Although the temporal pattern of behavior of all relevant celestial bodies is impossibly intricate, prediction can be achieved by putting together the simple cycles for the elements involved—the cycles of the sun, the moon, and the earth. In fact, if several interacting events have cycles of period p_1, p_2, and p_3 respectively, the composite cycle generated by all of these will have a period whose magnitude is the least common multiple of the component periods. Assuming that the difficulty of recognizing and retaining sequences increases very sharply with the length of the sequences, there is a considerable economy involved

* Received for publication April 23, 1968. The study reported was in part supported by Grant GS-1576 with the National Science Foundation. The author is indebted to David Christainsen, Jeffrey Brewer, and Donald Isaacs for able assistance in the administration of the studies and analysis of the data.

[1] A. H. Hawley, *Human Ecology: A Theory of Community Structure*, 1950.

in the more analytic approach. *(2) Cueing.* If the natural process is characterized by several distinct cycles, or by cycles of different lengths, telltale signs may indicate which one obtains on any occasion. Differential diagnosis of disease, for example, may hinge on the subtle divergence of symptom sequences at certain critical points. Stock-market experts depend on certain extramarket indicants, *e.g.* the volume of free disposable capital, to differentiate between otherwise similar market gyrations. *(3) Supercycles.* Even though the recurrence between successive event sequences is not exact, there may be a relation between the successive cycles which itself follows a definite pattern. Here again, the clearest example is astronomical: the gradual change of the daily cycle with the seasons.

Such hypothetical aids to cyclic identification will rarely be as clear-cut as these selected examples suggest, nor is it known how much they may be appreciated and used under natural circumstances. The purpose of this study was to translate the problem of cycle recognition into a laboratory context, to develop a paradigm for investigating the ability of naïve human *S*s to detect cycles and cyclic variations, and to explore the effects of the potential aids to recognition listed above as well as obvious quantitative parameters.[2] The complexity of the problem dictates that it be studied in highly schematic form, and it has been here cast in the guise of an anticipatory learning study with certain necessary modifications.

Method

The general task, as in conventional serial-learning studies, was one of predicting sequences of nonsense syllables. The chief departure from the standard situation was that the sequence varied on successive occasions, *i.e.* that the same set of nonsense syllables (simulated events) occurred in varying orders on successive trials. The *S*'s task, then, was to attempt to learn the several orders well enough so that when a few elements in a sequence were given, he could decide which sequence was in effect and predict accordingly. A secondary modification of typical anticipatory-learning studies was that the sequences began at a different point each time they were presented. This latter modification was based on methodological considerations rather than faithfulness to nature.

The three studies reported were conducted by different *E* assistants at about six-month intervals. Apart from certain details described below, the general design and administration of the studies were uniform enough so that they can be regarded as quasi replications.

Experimental design. The experimental treatments were defined by four two-level factors: *(1) Compound vs. scrambled sequences.* The sequences

[2] For a quite different approach to a similar problem, see C. Sengstake, Perception of deviations in repetitive patterns, *J. exp. Psychol.*, 70, 1965, 210–217.

each consisted of 12 two-syllable items generated by three distinct initial CVCs and four distinct final CVCs. These items were used to make up two types of sequences as illustrated in Table I. The compound sequences were based on two joint cyclic progressions of the three-element initial set (BOP, ZAM, YAK) and the four-element terminal set (RIK, JIR, DAF, NEY). The former repeated four times and the latter three times in the 12-element sequence. Four distinct compound cycles were constructed by factorial combinations of two cyclic orders for the initial and terminal sets. The scrambled sequences were based on the same set of 12 double CVCs, but arranged in random order. *(2) Two vs. four sequences.* In each experimental session there were 16 trials consisting of eight repetitions each of two distinct sequences, or four repetitions of four distinct sequences. All sequences, of course, consisted of the same 12 two-syllable items, but in different order. *(3) High vs. low transition redundancy.* In the first two replications, a high-redundancy condition was obtained by repeating each sequence with a probability of .5; in the low-redundancy condition this probability was .25. In the third replication, the high redundancy was further increased by changing the sequences in regular order; the low-redundancy transitions were completely random. *(4) Cue vs. no-cue to sequence at transition.* Under the cue conditions, a card inserted between each pair of sequences, *i.e.* at each sequence-transition point, provided a code letter for the forthcoming sequence. A blank card was inserted to indicate transition points in the no-cue conditions.

Procedure. In two replications, *S*s were run two at a time for administrative economy; in another replication they were run individually. Other changes in procedure were motivated by an effort to increase the sensitivity of the design in view of great individual differences in task performance. In the first replication, two *S*s were run in each of the 16 factorial conditions. In the second replication, each *S* was run in two sessions: in the second session, *S* was assigned to the opposite level on each of the four main factors from the level in which he had performed on the first session.[3] In the third replication, as in the first, two

TABLE I

Illustrative Form of Stimulus Sequence

	Compound				Scrambled	
Step				Step		
1.	BOP	NEY		1.	BOP	JIR
2.	ZAM	RIK		2.	ZAM	NEY
3.	YAK	JIR		3.	YAK	NEY
4.	BOP	DAF		4.	YAK	JIR
5.	ZAM	NEY		5.	ZAM	RIK
6.	YAK	RIK		6.	BOP	RIK
7.	BOP	JIR		7.	ZAM	JIR
8.	ZAM	DAF		8.	YAK	DAF
9.	YAK	NEY		9.	BOP	NEY
10.	BOP	RIK		10.	ZAM	DAF
11.	ZAM	JIR		11.	BOP	DAF
12.	YAK	DAF		12.	YAK	RIK

[3] Thus the *S* who was in the compound, two-sequence, high-redundancy, cue condition in the first session changed to the scrambled, four-cycle, low-redundancy, no-cue condition in the second session. This design provided an estimate of mean effects and second-order interactions that was independent of *S* differences but confounded the latter with first- and third-order interactions.

Ss were run in each factorial condition, but these *Ss* were also given a preliminary test on a presumably comparable task, namely four successive learning trials with a set of 12 single CVCs repeated in the same order each time but with different starting points. The twofold rationale of this warm-up task was first to stabilize performance (which it apparently did) and to provide a covariate (which it did not).

All *Ss* were given written instructions, and any questions were answered except for those bearing on the basic construction of the sequences. The *Ss* were provided with a tray containing stacks of each double CVC printed on IBM cards. The stimulus sequences were also presented by IBM cards in a stack. A pink Sequence Change card (which contained the cue in the cue conditions only) was presented first, then 12 cards of the sequence, then another Sequence Change card and a new sequence. Before turning each stimulus card, *S* made his prediction by selecting a response card from his tray and placing it in a pile facedown. After all 16 trials, the *Ss* were queried as to their recognition of the sequence construction, orally in the first two replications and on a written questionnaire for the third. The stimulus and response cards were kept in order as they were stacked and were later scored by a computer program.

The *Ss* were Tufts University undergraduates, men and women. In each of the three studies *Ss* were paid at an hourly rate and offered a $5.00 bonus for the best performance. The administration, including instructions and the post-session questionnaire, lasted 1–1½ hr.

Results

Individual analyses of variance for each of the three replications yielded almost identical results. Accordingly, a consolidated analysis, based on the treatment means for the three replications, is presented as Table II. The overall effect of treatments was highly significant, but this is attributable entirely to the Compound vs. Scrambled Sequence variable. The other three variables and their interactions failed to have a significant effect in this table and in the individual analyses. The mean proportion of correct responses was .30 for the compound sequences and .15 for the scrambled sequences; sheer chance is .083.

TABLE II

Analysis of Variance

Source	*df*	*MS*	*F*
Treatments	15	249.8	3.61**
Compound vs. scrambled sequences	1	2509.0	30.48**
Two vs. four sequences	1	111.0	
High vs. low transition redundancy	1	150.0	
Cue vs. no-cue to sequence at transition	1	72.0	
Pooled interactions	11	82.3	
Studies	2	224.5	3.26*
Conditions × Studies	30	69.1	

* $.05 < p < .10$.
** $p < .01$.

Turning to within-session changes in performance, both step effects (increases in accuracy across the 12 steps in each trial) and trial effects (increases in accuracy across the 16-sequence presentation) were significant for all three replications and followed a similar pattern. Consolidated analyses of variance indicated significant linear effects for both the step and trial variables, but in both cases there were also significant interactions between the linear component and the Compound vs. Scrambled variable (for the step effects, F [1, 10] $= 6.68$, $p < .05$; for the trial effects F [1, 14] $= 52.70$, $p < .01$). Inspection of the corresponding means indicates that *Ss* in the compound conditions 'caught on' to the pattern at about the third card in each trial, showing abruptly improved performance at that point. *Ss* in the scrambled conditions improved steadily but slowly across the 12 steps. Similarly the between-trial learning of *Ss* in the compound-sequence conditions was much more rapid than of those *Ss* in the scrambled-sequence conditions, although there are no evident discontinuities in either of these learning curves.

As the results indicate that only the Compound vs. Scrambled variable significantly affected performance, it is of interest to examine whether this effect depended on *Ss'* awareness of the sequence construction. Informal questioning of *Ss* following the final two quasi replications had indicated that there was very little awareness, and inspection of the questionnaires for the third replication bore this out. By the most generous interpretation, only two or three of the *Ss* in the compound-sequence conditions really 'cracked' the construction of the sequences. No *S* mentioned use of the sequence-redundancy or cue conditions.

Discussion

On the evidence of this study, the regular or random-interval construction of sequences is overwhelmingly the major factor in determining how well they are learned and recognized. Despite the uniformly negative results obtained in these task conditions, however, it is difficult to accept that cueing, regularity of sequence transition, and number of distinct sequences would not affect performance under other conditions. It is a plausible conjecture that other variables would become operative if a generally improved level of performance were achieved.

Among the modifications in procedure that might contribute to this objective would be, first, a reduction in sequence length. In

order to retain the method of compound-sequence structure employed in the present study, this would entail the use of sequences of six items, with two and three elements in the component sets. However, the compound-scrambled distinction may be regarded as sufficiently well established so that this variable might be dropped altogether, permitting sequences of arbitrary length. A second possibly facilitating modification would be to separate the sequence-learning and recognition stages, the question then being how quickly an *S* who had mastered several variants of a cycle could recognize and 'lock in' to the correct cycle. The cue condition in the present studies was, of course, designed to differentiate between the cycle recognition and reproduction processes, but it did not do so. One obvious way to strengthen the cue effect would be to associate a distinctive tag with each element in a cycle instead of inserting it between cycles.

The reduction of the effects of individual difference could be achieved most economically by having each *S* perform under more conditions. The decision not to attempt this in the present studies was based not on design but on humanitarian considerations, as *S*s found this a very taxing task. However, if the above suggestions for making the task easier were incorporated, and if the administration were automated, it should be feasible to obtain more data per *S*.

Among other modifications to broaden the base of generalization and to align the laboratory task more closely with the natural situation, performance might be investigated under conditions in which there is no marked break between cycles and in which there are varying numbers of successive items within a cycle before transition. Elements might be symbols instead of nonsense figures or, better, might be primitive "events" as in the Michotte studies.[4] Although these changes would probably make the task more difficult, they would also tend to make it more appealing. In any case, the laboratory paradigm explored in these studies seems to offer a point of departure for investigating an extremely important class of human conceptual activities.

[4] H. E. Michotte, *La perception de la causalite*, 1946.

NOTES AND DISCUSSION

ESTIMATES OF LENGTH IN A MODIFIED MÜLLER-LYER FIGURE

Fisher has recently suggested that a new insight can be gained into the causes of the Müller-Lyer illusion by rejecting the traditional emphasis upon the effects of ingoing and outgoing arrowheads.[1] Instead, he argues, the interaction between the two sides of the figures should be examined, so that the larger of the two illusion figures is considered as a horizontal line bounded by two pairs of lines converging upon it, and the smaller figure as a line bounded by two pairs of lines converging away from it. The illusion can then be accounted for on the principle that "horizontal lines more closely adjacent to the locus of intersection of inclined lines appear to be increasingly longer than those which are situated further from it."[2] This principle can be applied to a wide variety of "angle illusions" and has the advantage of revealing a close connection between the Müller-Lyer and Ponzo illusions.

Unfortunately, this elegant simplification is demonstrably false. Fig. 1 shows the Müller-Lyer figures with the "locus of intersection of inclined lines" (O) drawn by producing the lines AX, BY. It is apparent from elementary geometry that the perpendicular distance (OC) of the horizontal line (AB) from O is constant,

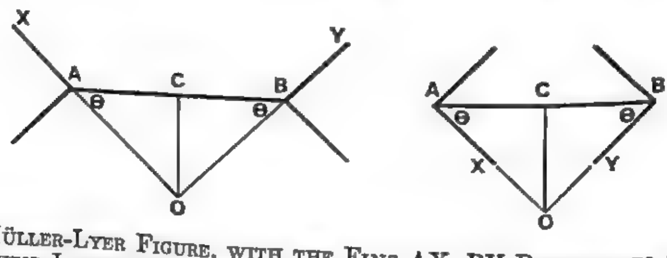


FIG. 1. MÜLLER-LYER FIGURE, WITH THE FINS AX, BY PRODUCED TO THE POINT OF THEIR INTERSECTION WITH OC, THE PERPENDICULAR BISECTOR OF AB

* Received for publication January 23, 1969. The author wishes to thank Mr. A. J. Watson for helping to draw his attention to the argument summarized in Fig. 1; and Professor O. L. Zangwill for his encouragement of the work reported here.

[1] G. H. Fisher, A common principle relating to the Müller-Lyer and Ponzo illusions, this JOURNAL, 80, 1967, 626–631.

[2] Fisher, *op. cit.*, 629.

being equal in each figure to Tan $\Theta \times AB / 2$, where Θ is half the angle included in the arrowhead. It is not the case, then, that the distance of the horizontal line from the locus of intersection of inclined lines differs between the two Müller-Lyer figures, and Fisher's attempt to derive the illusion from the Ponzo is thus a failure.

It is possible that this objection arises from a misunderstanding of Fisher's proposal; but a further argument may be raised against any form of hypothesis seeking to explain the Müller-Lyer illusion by lateral interactions between the two arrowheads. The objection arises from an illusion (Fig. 2) much more obviously related to the Müller-Lyer than is the Ponzo figure. In this illusion, mentioned in one of its forms by Köhler,[3] expansion of the horizontal line occurs in the region of the arrowhead with outgoing fins, just as it does in the Müller-Lyer figure, and the effect is to move the subjective midpoint of the line away from the ingoing arrowhead. The objective midpoint (m) is thus seen as displaced towards the ingoing arrowhead. The illusion is extremely persuasive, but it is clear that no account of it can be given by Fisher's hypothesis, for the pairs of lines on each side of the horizontal are neither converging nor diverging; and even if they were, the resulting distortion should apply impartially to the line as a whole rather than to any part of it.

It is also worth mentioning that this illusion raises considerable difficulties for Gregory's hypothesis that the Müller-Lyer illusion is caused by misplaced constancy scaling.[4] The figure as a whole has two equally likely depth interpretations, so that the horizontal

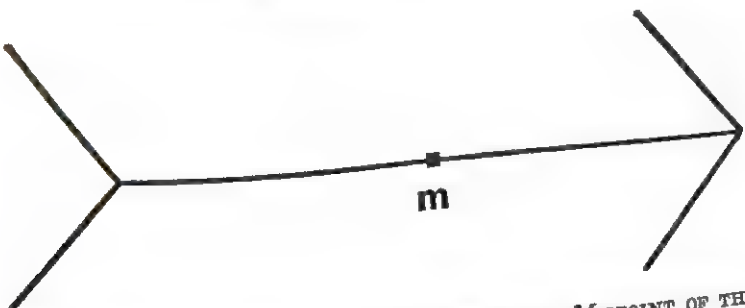


FIG. 2. MODIFIED MÜLLER-LYER FIGURE, WITH THE MIDPOINT OF THE HORIZONTAL LINE AT m

[3] W. Köhler, *Gestalt Psychology*, 1947, 170.
[4] R. L. Gregory, Distortion of visual space as inappropriate constancy scaling, *Nature*, 199, 1963, 679–680.

line can be considered as sloping away from the observer either in the direction of the ingoing or of the outgoing arrowheads. In the latter case, the direction of the illusion is correctly predicted by considerations of constancy scaling, but there is no evidence to suggest that the illusion reverses in direction when the alternative depth interpretation is made by the observer. This, of course, does not refute Gregory's theory, for he insists that the direction of an illusion has nothing to do with apparent depth as such. It is logically possible to argue that the "primary scaling" mechanism favors one of the two perspective interpretations, but this would further support the claim of Humphrey and Morgan that a theory appealing to a hidden constancy mechanism manifested only in the illusions themselves is essentially untestable.[5]

Some preliminary observations on the figure raise interesting questions about the nature of subjective measurement. *S*s ($N = 10$) were instructed to make two marks on a 'neutral' horizontal line such that they saw the distance between them as equal to *half* the length of a 9-cm. horizontal line bounded by ingoing and outgoing arrowheads as in Fig. 2. The objective midpoint of the line was not marked, and each arrowhead consisted of two 2.85-cm. fins enclosing an angle of 45°. The 'neutral' line, 20 cm. long, was drawn on a separate sheet of paper placed either to the right or the left of the illusion figure depending on whether *S* was right- or left-handed respectively; and was an extension of the horizontal line in the illusion figure except for a gap of 4 cm. between them. The illusion figure itself was placed with either the ingoing or the outgoing arrowhead on the side of the 'neutral' line, each *S* making an estimate in both conditions on lines presented on separate sheets of paper. *S*s were also asked to mark off on 'neutral' lines a distance seen as equal to the *whole* length of the horizontal line in the illusion figure, again making separate judgments in the two orientations of the figure. Thus four judgments were made by each *S*, in an order varying randomly from *S* to *S*.

There was no apparent effect of the orientation of the figure upon judgments, so readings for the two orientations were averaged for each *S* to obtain estimates of the subjective half and whole lengths of the line. Seven *S*s turned out to make an *overestimation* of the half length (average error = + .25 cm.), and this effect was significant at the .05 level (two-tailed Wilcoxon *t* test).[6] This would

[5] N. K. Humphrey and M. J. Morgan, Constancy and the geometric illusions, *Nature*, 206, 1965, 744–745.

[6] S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*, 1956, 75–83.

lead one to predict a large overestimation of the whole length, but in fact there was a nonsignificant effect in the direction of *underestimation* (average error = −.24 cm.). The surprising result is thus that the whole is considerably less than the sum of its parts. If each *S*'s estimate of the whole length is subtracted from twice the estimate of half length, eight *Ss* showed a positive discrepancy and only one a negative (average discrepancy = .72 cm.). The observed discrepancies had a probability of less than .01 (two-tailed) on the null hypothesis that the discrepancy between judged whole length and twice the judged half length would be in no consistent direction (Wilcoxon *t* test).

At first sight, the overestimation of half length may seem a direct consequence of the shift in subjective midpoint (Fig. 2), but in fact there is no logical connection between the phenomena. Consider the following fallacious argument. We know that the distance from the ingoing arrowhead to the subjective midpoint of the figure is objectively greater than the half length as measured by a ruler (Fig. 2). *Ss* copying this subjective half of the figure thus make an overestimation. But by definition, the other subjective half of the figure is equal to this and is also overestimated. Whichever half of the figure *Ss* attempt to copy, then, they make an overestimation. The fallacy here consists first in assuming that a certain physical distance in the illusion figure is *faithfully* copied by *S* and then in ignoring this axiom when considering the objectively smaller half figure. If we had considered the latter first and used the same process of reasoning, the conclusion would have been reached that *Ss* underestimate the half length. All attempts to derive the overestimation of half length as a *necessary* consequence of the shift in subjective midpoint seem likely to fall into this trap.

The main implication of these observations is that the judgment of the length of a bounded line is not related in any simple way to judgment of a segment of that line. The circumstances in which this is so may turn out to be either more general than our using an illusion figure might suggest, or highly specific to the testing conditions used in this investigation. In either case, the point is made that the human observer is capable of working with a geometry that he would undoubtedly repudiate if his attention were drawn to its axioms. The same point, incidently, can be made by Fig. 1. Although AB is subjectively different in length in the two diagrams, there seems to be little difference either

in Θ or in the length of CO; and this is obvious nonsense in Euclidean terms. Given that subjective geometry is patently not Euclidean, a promising approach to the illusions would be to investigate the possibility of the existence of a set of axioms in which these facts of perception no longer seem logically inconsistent.

University of Cambridge, England M. J. MORGAN

THE EFFECT OF MONOCULAR AND BINOCULAR OBSERVATION ON VISUAL ORIENTATION DURING HEAD TILT

Experiments concerned with binocular judgments of the visual vertical in the absence of a visual surround have demonstrated a high degree of accuracy when the head is upright.[1] However, differences occur between judgments made with the left and right eyes.[2] Using the left eye the visual vertical is displaced counterclockwise of the gravitational vertical, whereas clockwise displacements are produced with the right eye. Schneider has demonstrated that left-eye-dominant *Os* (as determined by a sighting test) perform binocularly as they do with the left eye only and that right-eye-dominant *Os* have a visual vertical similar for binocular and right-monocular viewing.[3] Although this difference has been clearly shown for verticality judgments with the head upright, little investigation has been directed to the question of whether this difference is main-

* Dr. Wade, the author of this note, carried out the research at Monash University, Melbourne, Australia.

[1] C. W. Mann, N. H. Berthelot-Berry, and H. J. Dauterive, The perception of the vertical: I, Visual and non-labyrinthine cues, *J. exp. Psychol.*, 39, 1949, 538–547; E. Neal, Visual localization of the vertical, this JOURNAL, 37, 1926, 287–291; D. C. O'Connell, R. G. Lathrop, D. J. Weintraub, and T. J. McHale, Apparent verticality: Psychophysical error versus sensory-tonic theory, *J. exp. Psychol.*, 73, 1967, 347–353; N. J. Wade, Visual orientation during and after lateral head, body, and trunk tilt, *Percept. & Psychophys.*, 3, 1968, 215–219; N. J. Wade, The effect of stimulus line variations on visual orientation with head upright and tilted, *Aust. J. Psychol.*, in press; D. J. Weintraub, D. C. O'Connell, and T. J. McHale, Apparent verticality: Fundamental variables of sensory-tonic theory reinvestigated, *J. exp. Psychol.*, 68, 1964, 550–554.

[2] M. H. Fischer, Messende Untersuchungen über die Gegenrollung der Augen und die Lokalisation der scheinbaren Vertikalen bei seitlicher Neigung (des Kopfes, des Stammes und des Gesamtkörpers): I, Mitteilung; Neigung bis zu 40°, *V. Graefes Arch. Ophthal.*, 118, 1927, 633–680; B. O. Peirce, The perception of horizontal and vertical lines, *Science*, 10, 1899, 425–430; C. W. Schneider, Monocular and binocular perception of verticality and the relationship of ocular dominance, this JOURNAL, 79, 1966, 632–636.

[3] Schneider, *loc. cit.*

tained for judgments during tilt. There is a suggestion from previous research that this may be the case.[4] The purpose of the present experiment was to investigate visual orientation with left- and right-monocular and binocular observation in various head positions. A range of head tilts was employed so that differences in the trend of judgments under the various viewing conditions could be tested.

Method

The apparatus for controlling head position consisted of side head clamps and an individual dental composition bite board. These could be rotated to achieve lateral head tilt, the degree of which could be read to the nearest ¼° by means of a protractor scale. The line of light (1.94 ftL.) corresponded to an aperture (152.0 mm. long and 0.9 mm. wide) in an otherwise lighttight lamp housing, 183 cm. from *O*'s eyes. The line could be rotated frontoparallel to *O* with its center in the median plane and adjustable to eye level. The orientation of the line, which could be read to the nearest 0.1°, could be controlled either by *O*, using a knob connected to a mechanical drive system, or by *E*. The apparatus was situated in a dark room.

Thirty *O*s from an introductory course in psychology participated. They were assigned randomly to one of three groups (left- and right-monocular and binocular observation), 10 *O*s in each group. First, *O* made 10 adjustments of the line to the visual vertical with head upright from starting positions of 5°, 15°, 30°, 60°, and 80° clockwise and counterclockwise, presented in random order. Then, visual-vertical judgments were made from head tilts of 10°, 20°, 30°, and 40° left and right from starting positions of 15° and 30° clockwise and counterclockwise. The head was returned to the upright after each of the 32 adjustments, which were presented in random order. Finally, a block of 10 more adjustments was made with the head upright, following the same procedure as the series prior to head tilt. Throughout the experiment the *O*'s eye(s) were open only during an adjustment and closed at all other times.

Results

The data for the visual vertical with head upright were analyzed separately. The influences of starting position and block order were not significant ($p > .05$), and therefore the visual vertical for each *O* was averaged over these. Denoting clockwise deviations of the visual from the gravitational vertical as + and counterclockwise as −, the group means were −1.50° for the left eye, +0.40° for the right eye, and +0.35° for both eyes. A one-way analysis of variance yielded a significant effect for the different observation conditions ($p < .001$). Post hoc tests for the differences between the means were carried out using the *R* technique.[5] The visual vertical with

[4] Fischer, *loc. cit.*; Peirce, *loc. cit.*

[5] R. S. Rodger, *Intermediate Statistics*, 1965, 41–51; R. S. Rodger, Type I errors and their decision basis, *Brit. J. math. stat. Psychol.*, 20, 1967, 51–62.

the left eye differed significantly from that with the right eye ($p < .01$), and the algebraic mean of the left- and right-eye conditions was significantly different from the binocular condition ($p < .05$).

The visual vertical was averaged over starting positions for each tilt, and these values are given in Fig. 1 for all observation conditions. The trends for judgments over head tilts were analyzed by Grant's model.[6] The means for the head-upright condition were included in this analysis. For the three curves taken together, the significant trend components were linear ($p < .001$), quadratic ($p < .05$), and cubic ($p < .001$). However, there were no significant differences between the overall means or the trends for the three observation conditions ($p > .05$).

Discussion

Visual-verticality judgments made with the head upright were significantly affected by the eye(s) used for viewing the line of light. Judgments with the left eye were counterclockwise of those with the right eye. This result is in accord with the findings of both Peirce and Schneider.[7] The difference between the left- and right-monocular judgments was 1.9°, whereas in Schneider's ex-

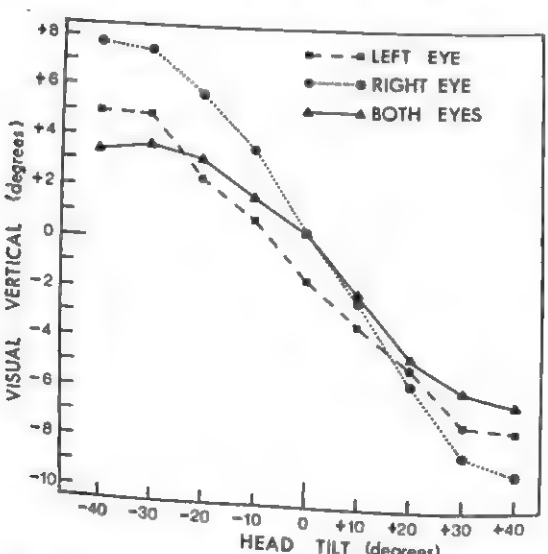


Fig. 1. Visual Vertical as a Function of Head Tilt under Conditions of Left- and Right-Monocular and Binocular Observation

[6] D. A. Grant, Analysis-of-variance tests and the analysis and comparison of curves, *Psychol. Bull.*, 53, 1956, 141–154.

[7] Peirce, *loc. cit.*; Schneider, *loc. cit.*

periment it was approximately 1.5° (when the different groups are compared). The binocular observation group performed similarly to those using the right eye only, and the mean for the former differed significantly from the average of the left- and right-monocular conditions. In view of Schneider's result concerning ocular dominance, it is possible that the binocular group contained mainly right-eye-dominant *Os*.

During head tilt the visual vertical was located on the opposite side of the gravitational vertical to tilt. This is referred to as the E effect, and its converse is called the A effect.[8] The E effect occurred for all tilts under all viewing conditions; it increased up to 30° and then tended to level off. For all conditions taken together, the trend was expressed in terms of the linear, quadratic, and cubic components, the linear and cubic accounting for most of the variance. The mean E effect for the right-eye observation group was greater than that for the left-eye group, and both were greater than the E effect with binocular observation at large tilt angles (30° and 40°). However, due to the large individual differences in this task, the differences between the trends for the various observation conditions were not significant. Moreover, the overall means did not differ significantly. That is, the difference between the conditions found with head upright was not maintained for judgments during tilt. The visual vertical with the right eye was clockwise of that with the left when the head was upright, but this difference was reversed with high degrees of right tilt.

Fischer investigated the effect of left- and right-monocular observation on the visual vertical with head tilt.[9] The two *Os* tested showed a monocular difference similar to that here found for the upright posture, and this was approximately maintained during head tilt. However, it should be noted that the judgments of these two *Os* during head tilt would be considered atypical in comparison to the *Os* tested in the present experiment. One *O* produced the A effect for all tilts, and the other gave the E effect for one direction of tilt and the A effect for the other. Only one *O* of the 30 here tested gave the A effect for all directions of tilt; he was in the binocular group, and this resulted in a reduction in the values of the curve for this condition. This point apart, similar directional

[8] H. Aubert, Eine scheinbare bedeutende Drehung von Objekten bei Neigung des Kopfes nach rechts oder links. *Arch. path. Anat. Physiol.*, 20, 1861, 381–393; G. E. Müller, Ueber das Aubertsche Phänomen. *Z. Psychol. Physiol. Sinnesorg.*, 49, 1916, 109–244.

[9] Fischer, *loc. cit.*

differences due to monocular observation would be expected irrespective of O's mode of response. The negative findings concerning the overall means for the three conditions suggests that the monocular differences found with head upright no longer hold for judgments during tilt. In addition, the sparse data obtained for judgments during body tilt do not indicate any systematic differences in visual orientation with the left and right eye.[10] Nonetheless, the variances associated with judgments during tilt were much greater than those with the head upright, and this factor, together with the use of different *Os* in the three conditions, may have acted to mask slight effects of the various viewing conditions.

It is possible that the left-right monocular difference with the head upright may have its basis in different ocular postures during such observation. That is, if ocular extorsion occurs under both conditions, the monocular visual vertical may represent the orientation of the retinal meridian. However, differences in monocular posture with head upright have not been investigated. Measurements of ocular countertorsion during tilt use the head-upright eye position as the reference against which other eye positions are compared. This procedure precludes the possibility of detecting monocular differences, and thus the hypothesis remains to be tested. Slight differences in ocular countertorsion between the left and right eye have been reported during head and body tilts.[11] In conclusion, it is considered that visual orientation during head tilt is not influenced by the eye(s) used for observation, unlike that for judgments with head upright.

Summary

The influence of left- and right-monocular and binocular observation on the visual vertical was investigated for head tilts between 40° left and 40° right. With the head upright the visual vertical for the left eye was counterclockwise of that with the right eye, confirming previous findings. During head tilt the visual vertical was located on the opposite side of the gravitational vertical to tilt (E effect), but the magnitude of the effect was not significantly influenced by the observation conditions.

Max-Planck-Institut für Verhaltensphysiologie N. J. Wade
8131 Seewiesen, Germany

[10] Fischer, *op. cit.*, 664; Peirce, *loc. cit.*
[11] Fischer, *op. cit.*, 633–680.

A NEW PHOTOGRAPHIC ILLUSION

Well-known to psychologists are a few photographically obtained figures characterized by certain gradients of light and shade which will, when viewed upside down, induce reversed perceptions of concavity and convexity.[1] Also well-known are line drawings of figures which, no matter whether viewed right side up or upside down, induce fluctuating perceptions of concavity and convexity.[2] The literature on perception does not, however, appear to include a photographic example of the latter type of ambiguous figure. This note describes such a photoillusion. The picture, shown in Fig. 1, was obtained by use of flash in a dark stair tunnel slanting downward between landings at an angle of about 40°. Since the roof is more easily perceived as an inverted, truncated cone than as a concave arch, the picture has been titled 'The Falling Arch of Fort Knox.'

Method

A lantern-slide projection was shown to 114 undergraduates of both sexes in four laboratory sections of an introductory course in psychology. The image was 42 in. high × 55 in. wide, with the nearest *S* 10 ft. away and the farthest about 33 ft. Before the showing of the picture, *Ss* were told that they would be shown a picture of some brickwork and two people, and to pay attention to the brickwork only. These instructions, read aloud by *E*, were mimeographed on a response sheet provided to each *S*. The response sheet also listed three questions that the *Ss* were to answer after having looked at the picture for 2 min. The first question asked whether the first appearance of the brickwork was as an arch, or as a cone or funnel. The second question asked which appearance of the brickwork, arch or cone, predominated with continued viewing. The third question asked whether the brickwork did or did not fluctuate between arch and cone with continued viewing. To eliminate questions as to the difference in appearance between an arch and a cone, *E* demonstrated the difference by curving a sheet of paper with his hands.

The 62 students in two of the sections, constituting Group RU, were shown the picture right side up for 2 min. (Test 1). After answering the questions, these *Ss* were shown the same picture upside down for 2 min. (Test 2), and again answered the same questions (repeated on the lower halves of their response sheets). Finally, the *Ss* were asked to look at the picture for 1 min., while it was presented upside down, to count the number of changes in the appearance of the brickwork (from arch to cone and vice versa) and to report the number on their response sheets. The 52 students in two other sections, constituting Group UD, were tested in the same way except that they saw

[1] N. Pastore, The influence of background on the perception of convexity and concavity, this Journal, 70, 1957, 131–132.

[2] An example is the "folded-paper" figure in F. L. Ruch, *Psychology and Life*, 6th ed., 1963, 296.

FIG. 1. THE 'FALLING ARCH OF FORT KNOX' PHOTOILLUSION

the picture presented first upside down (Test 1) and then right side up (Test 2). For the final count of fluctuations, the picture was right side up.

RESULTS AND SUMMARY

Chi-square tests applied to the frequency data of the two groups revealed no statistically significant differences between the groups on either Test 1 or Test 2 in their responses to the three questions and in numbers of fluctuations reported. Whether the picture was viewed right side up or upside down, then, did not significantly determine whether it was perceived as an arch or cone or whether it fluctuated. On Test I, of the total group of 114 *Ss*, the majority of *Ss* (59%) stated that their first impression of the brickwork was as an arch, while 41% saw it first as a cone; however, in predominant appearance 68% stated the brickwork appeared as a cone while 32% saw it as an arch. By contrast, on Test 2, for the majority of *Ss* the brickwork was a cone both on first impression (59%) and in predominant appearance (70%). Frequencies of *Ss* changing their first-impression responses from "arch" in Test 1 to "cone" in Test 2 were not significant in the RU Group but were

significant at less than the .01 level of confidence in the UD Group when evaluated with the McNemar test for the significance of changes.[3] There was, therefore, a somewhat greater initial tendency to perceive the brickwork as an arch when the picture was presented upside down than when it was shown right side up. On the final viewing of the picture, in the right-side-up position, the *Ss* in the UD Group showed a clearcut tendency to change their first-impression reports from "arch" to "cone" to correspond with their reports that the predominant appearance of the brickwork was as a cone during Test 1.

The distribution of numbers of fluctuations or changes from arch to cone and vice versa that occurred in the final 1-min. test was highly skewed to the right. In the total group the median number of fluctuations was 4.7, the mode was zero (30 *Ss*), and the range was from 0 to 36.

It may be stated, in conclusion, that perception of the brickwork as concave or convex did not depend greatly on whether the picture was viewed right side up or upside down; that with continued viewing, illusory appearance of the brickwork as a cone predominated; and that perception of the brickwork fluctuated to the same degree whether viewed right side up or upside down. This photoillusion, therefore, should be classed with the geometrical line drawings of ambiguous figures.

University of Maine

JOSEPH J. ANTONITIS

[3] S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*, 1956, 63–64.

GALILEAN AND NEWTONIAN INFLUENCES ON PSYCHOLOGICAL THOUGHT

The physical theory of Isaac Newton was presented to the world in the late 1680s under the title *Mathematical Principles of Natural Philosophy*. Among those whose interest it very soon attracted was a countryman of Newton's, a certain John Locke. Already, Locke maintained an avid interest in the relatively new "philosophy of mechanism," and he was quite conversant with the earlier mechanistic conceptions of such notables as Galileo and Robert Boyle. What he had heard of Newton's *Principia* intrigued him; but since

* This is a somewhat modified version of a paper presented at the 1968 meeting of the APA, San Francisco, California.

he did not have (as Newton put it) "good mathematical learning," he had to seek assurance from others that the mathematics it contained could be relied upon. Receiving such assurance, he then proceeded to read the *Principia* for its more general ideas—which he adopted. Several years later, this same John Locke published a work that effected a minor Newtonian revolution within psychology.

Indeed, it is not stretching the point too far to suggest that the whole of Locke's psychology can be seen as a kind of Newtonian cosmos in miniature. The particles or ultimate constituents of Locke's mental cosmos were, of course, the "simple ideas." They in turn could cohere to form aggregate or "complex" ideas. Knowledge, finally, was simply "the mind's perception" of whether ideas naturally, and "by some causes hitherto unknown, are either mutually impelled towards each other, and cohere to form regular figures, or are repelled and recede from each other."[1]

Was Locke consciously endeavoring to construct a Newtonian psychology? To answer this question, we should have to raise him from the dead; for while he lived, he neither affirmed nor denied it. In either case, though, this much is clear: (*a*) he was quite familiar with the Newtonian cosmology; (*b*) he accepted as true the refined scheme of mechanism that it embodied; and (*c*) in structure—though not, of course, in content—his psychology may be aligned almost point for point with the structure of the Newtonian cosmology. Conjoin with this the observation that Locke's *Essay* was written almost immediately after he became acquainted with the *Principia,* and you may safely judge that the correspondence was not entirely accidental.

The lesson to be gained from Locke's example is simply this: one need not be a physicist to be influenced by physical theory. Locke was not the first psychological thinker to be so influenced, and he most surely was not the last. What I wish to suggest in what follows is that the historical relations between physical and psychological theory have been more direct and intimate than we have customarily imagined.

Let me begin with a familiar example from pre-Lockean psy-

[1] This passage is not taken from Locke but from Newton's preface to the first edition of the *Principia* (I. Newton, *Mathematical Principles of Natural Philosophy*, A. Motte [trans.] and F. Cajori [ed.], 1934). Corresponding passages from J. Locke's *Essay* may be found in his important but often neglected chapter "On Knowledge in General," *An Essay concerning Human Understanding*, 2, 1894.

chology. René Descartes and Thomas Hobbes would have both no doubt resented being characterized as two peas out of the same psychological pod. And yet, their probable resentment notwithstanding, this is precisely what they were. For whatever their differences on specific points might have been, there was one important and far-reaching issue on which they both flatly agreed: namely, that the study of psychology is ultimately a study of *matter* ("animal spirits" or whatever) in *motion*. It of course goes without saying that this emphasis on matter-in-motion was not the novel creation of Descartes, or Hobbes, or indeed of any other psychological thinker of the period. Every age has its grand ideas, and the grand idea of the early and middle seventeenth century was just this: that matter and motion constitute the warp and woof of existence-at-large. Usually, this idea is referred to as 'the mechanical conception of nature' or, more simply, as 'mechanism.'

Although the mechanism to which Descartes and Hobbes adhered had already been expanded into a general cosmology, its birthplace and permanent residence was in physics. Medieval physics, its predecessor, had been a great tangle of form and substance, essence and attribute, and the like. In its own way, it explained just about everything, but so cumbersomely that, by the sixteenth century or so, it began to seem hardly worth the effort. Mechanism, on the other hand, promised to account for physical phenomena by way of a single, unitary principle: matter-in-motion. By the middle of the seventeenth century, at any rate, the general outlines of the new mechanism were fairly widely known. The whole of nature rather resembled a great game of billiards. Matter, which was all that it contained, was held to exist in the form of discrete corpuscles that could affect one another only by coming into direct contact. Thus, in the words of the new conception's foremost spokesman, Galileo, the world was now to be seen as consisting in "a multitude of minute particles having certain shapes and moving with certain velocities."

For our own purposes, the important thing about Galilean physical theory is that the psychologists of the earlier seventeenth century took it over in a wholesale fashion and constructed their psychologies on the basis that it provided. And thus, without so much as running a single subject, they arrived at several psychological conceptions that have haunted us right up to the present. I shall mention these only briefly, for I wish to press on to other matters.

(1) To the question of what is involved in an act of perception, there are essentially two kinds of answers: the first, that we perceive things *objectively;* the second, that we perceive things only *subjectively*. Descartes and Hobbes both argued that the truth of the matter lay on the side of subjectivity. This theory of the subjectivity of perception did not, of course, originate with Galilean physical theory, but it was a thoroughly Galilean notion nonetheless. Galileo himself had put it forward when he attempted to give a mechanical explanation of heat. Later, John Locke employed the same line of reasoning to arrive at his factitious distinction between primary and secondary sensory qualities: If matter-in-motion is all that *obj*ectively exists in the world, then clearly any perceptions involving anything other than matter-in-motion must be *sub*jective. The theory was a psychological one, to be sure; but, as Locke admitted, it was the result as well of "a little excursion into natural philosophy." In any case, it is fair to say that Descartes and Hobbes would not have arrived at the theory of the subjectivity of perception so explicity had it not been *demanded* by the Galilean cosmology.

(2) Descartes and Hobbes both considered the mind's perception of an object to be mediated by the nervous system, and the nervous system which they imagined to do this mediating was of a distinctly Galilean sort. Indeed, it was but the Galilean universe in microcosm: "a multitude of minute particles having certain shapes and moving with certain velocities." Descartes, of course, accepted the then-current notion of "animal spirits"; Hobbes spoke more neutrally of "some internal substance of the head"; but the effect in both cases was quite the same. The motion of external matter gives rise to internal motion in a sense organ; this motion, in turn, is transmitted by way of the nerves into the brain; and there it gives rise to a perception. Thus, in the well-known words of Hobbes, a percept "is but an apparition unto us of the motion, agitation, or alteration, which the [perceived] object worketh in the brain."[2]

Now, given this particular scheme of things, a certain other psychological inference naturally presented itself: If perception is but the mental counterpart of motions in the brain, then does it not follow that all other mental occurrences are but these *same* motions, attenuated and variously combined? The answer is that it does so follow—providing we assume (*a*) that *all* mental activity

[2] T. Hobbes, *Human Nature*, in W. Molesworth (ed.), *The English Works of Thomas Hobbes*, 4, 1840.

is dependent upon motion in the brain and (*b*) that *all* such motion derives ultimately from external sources. Descartes had precluded both of these provisions with his assumption of a *res cogitans*. Hobbes, however, was able to take them entirely for granted; and thus it was that he arrived at the same well-worn doctrine of psychological empiricism that was later elevated by Locke and his followers to the level of a first principle.

(3) The third Galilean influence that I shall mention is a particularly curious one. Some two millennia before Hobbes, Aristotle had advanced three principles to account for the association of ideas; these being the familiar "contiguity," "similarity," and "contrast." Hobbes's problem was that his motion-in-the-brain theory obliged him to reject all but the first of these principles. He could well imagine how motions in the brain might tend to cohere by virtue of their initial contiguity (for this was simply to say that motions that begin together remain together); but it was not at all clear how such motions could be brought together by reason of their similarity or contrast. Thus, as a result of his Galilean assumptions, Hobbes had to portray the whole of association as the result of a single principle: contiguity. In this direction, one could trace a fairly unbroken line extending from Hobbes, to Hartley, to James Mill, even on to some of our modern learning theorists.

Here, then, are three of the more far-reaching psychological consequences of Galilean physical theory: the theory of the subjectivity of perception, the doctrine of psychological empiricism, and the reliance upon contiguity as the sole principle underlying the association of ideas. Just in passing, allow me to mention one further consequence, namely, the theory that all behavior is essentially reflexive in character. As I have suggested elsewhere, the birth of this reflex model in psychology can be seen as a direct outgrowth of what Descartes and Hobbes assumed to be the Galilean properties of the nervous system.[3]

The Galilean period in the history of physics came abruptly to an end with the publication, in 1687, of Newton's *Principia*. The influence of this important work upon psychology may be assessed in a variety of ways, of course; but for the moment let us consider the effects of only two of its more novel and consequential features.

[3] R. Lowry, The reflex model in psychology: A genealogical sketch, paper read at the APA meeting, August, 1968.

(1) The first pertained to the issue of action at a distance. In Galilean physics, and indeed throughout the history of physical inquiry, it had been assumed that motion could be communicated from one body to another only by way of direct contact. Newton, on the other hand, demonstrated that physical phenomena could be accounted for more fully and precisely if it were allowed that action could also be communicated, from a distance, by way of attractive and repulsive *forces*.

From this, it did not take psychological thinkers long to imagine that the mind might be a kind of Newtonian cosmos in miniature, in which ideas get associated with one another according to whether they attract or repel. Locke, for example, suggested that ideas might have a natural "connexion and agreement" or "disagreement and repugnancy."[4] Hume, somewhat later, regarded the association of ideas as a "gentle force"—as "a kind of *attraction*, which in the mental world will be found to have as extraordinary effects as in the natural and to shew itself in as many and as varied forms."[5] Although it cannot be entered into just now, I think it can be shown that the attraction-repulsion metaphor never went much further than this among the English because it was eventually seen to be incompatible with the endemic doctrine of psychological empiricism. In any case, where the Newtonian metaphor really made its appearance was in nativistic, post-Kantian Germany. Herbart, for example, spoke explicity of repulsive forces and implicitly of attractive ones; and it was precisely on this basis that he arrived at germinal versions of such important psychological ideas as "repression," "threshold of consciousness," and "apperceptive mass."[6] Herbart willingly acknowledged his debt to Newtonian physical theory. As we shall see presently, he was not the last Germanic psychological thinker to contract such a debt.

(2) The second feature of the Newtonian system that we may consider was the conception of rectilinear inertia. Embodied as it was in Newton's three laws of motion, this conception implied a conservation of mechanical energy; and from this there evolved the idea that energy-in-general is conserved. Thus, the familiar 'law

[4] Locke, *op. cit.*, 2, 167.

[5] D. Hume, *A Treatise of Human Nature*, 1888, pp. 12–13.

[6] Herbart began with two assumptions: (1) that "opposed conceptions" repel one another in a degree proportional to their "mass," and (2) that all "conceptions" are attracted toward the "center" of consciousness in equal degree. From this, it was a fairly simple matter to work out a hypothetical scheme of mental "statics" and "dynamics." See J. F. Herbart, *Schriften zur Psychologie*, in G. Hartenstein (ed.), *Sämmtliche Werke*, 6–7, 1886–1889.

of the conservation of energy': the energy within any closed physical system remains *quantitatively* constant, although it may undergo various *qualitative* transformations.

What is often forgotten about Hermann von Helmholtz, the framer of this 'law,' is that his principal interest in physical theory lay in its application to organic functioning. He was, indeed, at the forefront of the so-called German Biophysical Movement, the expressed aim of which was to demonstrate that all forces active within the living organism are (in the words of Ernst Brücke and Emil Du Bois-Reymond) "equal in dignity to the chemical-physical forces inherent in matter [and] reducible to the force of attraction and repulsion."[7] And it was with this end clearly in view that Helmholtz prepared his now-famous paper "On the Conservation of Force."[8] Thus, on the basis of his so-called constancy principle, Helmholtz and his followers could now see organic functioning as a complex but nonetheless mechanical process of energy exchange and transformation. In the living organism, as in the universe-at-large, the various forms of energy are converted into one another with neither a loss nor a gain in overall quantity. The only difference between the two is this: The universe is an isolated physical system; thus, it receives no energy from other systems, nor does it lose any to them. The organism, on the other hand, is only relatively isolated; it receives energy from, and expends energy upon, the environment.

There was also a more specific implication of the Helmholtzian model. The activity of the nervous system was by this time recognized to be electrical in nature; and on the analogy provided by known electrical circuitry, it was quite naturally assumed that the nervous system *as a whole* functions in the manner of a simple electrical conductor. Thus, according to this theory, no matter what form of energy it is that excites a sense organ, it will in any case be converted into electrical energy. In this form, it will then be conducted through the nervous system, to an effector, where it will again be *qualitatively* transformed. According to the constancy principle, moreover, the *quantity* of energy will remain unaltered throughout all of these transformations. Thus, we may expect the magnitudes of stimulus and response to be, if not identical, then at least proportional.

[7] E. Du Bois-Reymond, *Zwei grosse Naturforscher des 19. Jahrhunderts: Ein Briefwechsel zwischen Emil Du Bois-Reymond und Karl Ludwig*, 1927.

[8] H. L. F. von Helmholtz, *Ueber die Erhaltung der Kraft*, 1847.

Helmholtz' paper "On the Conservation of Force" was published in 1847. Three years later, in 1850, Gustav Theodor Fechner had his famous revelation. Whether the first caused the second, we can not say; but it is in any case clear that Fechner's psychophysics of 1860 took the constancy model of the nervous system entirely for granted. Without it, indeed, Fechner would never have seen the connection between his *Tagesansicht* and the measurement of "outer" psychophysical relationships. For as he observed in the opening pages of his *Elemente*, the mediated relationships of stimulus and sensation "fulfill completely the concept of a functional relationship only under the supposition that the mediation [*i.e.* the neural conductor] enters into the relationship."[9] Fechner, in fact, was so taken with the Helmholtzian constancy principle that he devoted almost an entire chapter of his *Elemente* to its exposition. He conceded that the applicability of the principle to psychophysical processes had "not yet been demonstrated." But it can scarcely be doubted, he continued, "that all experiences . . . agree with this law and can be interpreted without trouble only by means of it."[10] It is also fairly well known that the great Helmholtz responded in kind, hailing Fechner's psychophysics as an important step along the road to a truly scientific—by which he meant physicalistic—psychology.

Fechner's psychophysics was the first psychological issue of the Helmholtzian constancy principle; but it was not, I think, the most far-reaching. For this, the credit must go to Freud. Freud's contact with the so-called Helmholtzian school of physiology has been well chronicled elsewhere and need not be entered into here; suffice it to say that he imbibed deeply of the teachings and disciplines of this school during his earlier years. If you look into his "Project for a Scientific Psychology" of 1895, you will find an enthusiastic but abortive effort to construct a Newtonian-Helmholtzian psychophysiology.[11] If you look into his psychoanalytic theory of 1900, and thereafter, you will find an equally thorough though less explicit attempt to construct a Newtonian-Helmholtzian

[9] G. T. Fechner, *Elements of Psychophysics*, 1, D. H. Howes and E. G. Boring (eds.) and H. E. Adler (trans.), 1966, 8–9.

[10] Fechner, *op. cit.*, 1, 31.

[11] This was a portion of Freud's correspondence—published only after his death—with the Berlin physician Wilhelm Fliess: see S. Freud, *Aus den Anfängen der Psychoanalyse*, M. Bonaparte, A. Freud, and E. Kris (eds.), 1950, 379–466; and S. Freud, *The Origins of Psychoanalysis*, M. Bonaparte, A. Freud, and E. Kris (eds.), 1954, 355–445. "Entwurf einer Psychologie" was the German title accorded by the editors of the Fliess correspondence to this originally untitled draft.

psychology. At the heart of Freud's earlier "Project" was what he spoke of as "the principle of neuronic inertia," which was simply the constancy model of the nervous system in cytological guise. At the heart of his later theory was a kind of psychical inertia, which was the same model in psychological guise. In both cases, the underlying assumption was quite the same: The principal function of the "psychical apparatus" is to maintain a constant level of energy or "tension"; thus, its "primary process" consists in the immediate "discharge" of any excess "tension." I think it could be argued that any such homeostatic model of psychological functioning is essentially Newtonian-Helmholtzian in character.

I shall close this section with one further observation on Freud. In his introduction to "The Conservation of Force," Helmholtz wrote that "finally . . . we discover the problem of physical natural science to be, to refer natural phenomena back to unchangeable attractive and repulsive forces. . . . The solvability of this problem," he dilated, "is the condition of the complete comprehensibility of nature."[12] I think it can be shown that Freud—good physicalist that he was—spent the major portion of his own theoretical career attempting to solve this problem of "physical natural science" in the sphere of psychology. Consider, for example, the dynamics of those two conceptual bulwarks of psychoanalytic theory, "repression" and "displacement." By Freud's own admission, what is involved here is a complex of attractive and repulsive forces, not unlike those posited by Herbart nearly a century earlier. More to the point, consider his later theory of instincts: as Freud himself observed in one of his last written works, "the analogy of our two basic instincts ["life" and "death"] extends from the sphere of living things to the pair of opposing forces—attraction and repulsion—which rule in the inorganic world."[13]

Here, then, are a few of the more conspicuous Galilean and Newtonian influences upon psychological thought. Others could be cited, but these I think will suffice to make the point. I should now like to conclude with an observation upon the larger issue that is involved in all of this. The psychology of the medieval world is often said to have served as *ancilla theologiae*, the hand-

[12] Helmholtz, *op. cit.*, vii.

[13] S. Freud, An outline of psycho-analysis, in J. Strachey (ed.), *Standard Edition of the Works of Sigmund Freud*, 23, 1964, 149. The present writer has considered Freud's "physicalism" elsewhere at greater length; see R. Lowry, Psychoanalysis and the philosophy of physicalism, *J. Hist. behav. Sci.*, 3, 1967, 156–167.

maiden of theology. The psychology of the postmedieval world, on the other hand, may be said without too much exaggeration to have served as *ancilla physicae,* the handmaiden of physics. In both cases, what we have is a psychology that is modeled upon, and draws its major ideas from, the Grand Scheme of Things that happens to be in fashion at the time.

Whether this submission to the allure of the Grand Scheme of physics is good or bad, is a question not easily answered. Certainly, the science of chemistry was much advanced by the Newtonizing efforts of Lavoisier and Dalton; and it may be argued that psychology would do well to follow the same example—and hope for the best. Nor need the argument be restricted to 'Newtonian' physical models; it could be readily extended to some version of modern physical theory, or to some modern physical subtheory, such as has been haltingly attempted by the Gestalt and field theorists. For my own part, though, I must confess that I am less sanguine about the matter. Psychologists have been dabbling in physicalistic models now for the better part of three centuries; and I simply do not see that the models have repaid the effort invested in them. Perhaps they will begin to pay off in the future; but then again, perhaps they will not. In any case, we should do well to remain ever mindful of *other* possibilities and to remember that in all events the burden of proof falls squarely upon those who would employ such models.

Vassar College

RICHARD LOWRY

NEWTON, PHYSICS, AND THE PSYCHOLOGY OF THE NINETEENTH CENTURY

That psychologists, no less than any other kind of people, are prisoners of their culture is perhaps a commonplace. How they may escape being time-bound and space-bound is a less easy question. My interest in examining one small area in the history of psychology arises from my belief that answers to certain very critical questions have been overlooked and that if the issues are seen more clearly in historical perspective, the message will become more clear.

My main thesis is that the essence of Gestalt psychology, contrary to popular belief and to the statements of many eminent psychologists, lies in its insistence upon a monistic approach to psychologi-

cal fact—upon a kind of physicalism, if you will, and a rejection once and for all of dualistic premises, however deceptive may be the packaging. That most early psychologists of the introspectionist stripe were blatant dualists is readily recognized. It is my contention that most behaviorists are quite as much dualists because the concepts and explanations they use are only thinly veiled transformations of the spiritualistic explanations employed by the introspectionists. Is there any formal difference between habit strength and associative bonds? Between physiological need and desire? Admittedly, the linguistic style, the jargon, is changed but the substance remains much the same.

In pursuit of my thesis I had occasion to hunt for several quotations used by Köhler from older psychologists.[1] The results were quite startling. In Maxwell, in Planck, in Mach, one finds views that leave the way open for bridges among the natural sciences. These views were almost completely ignored by psychologists. At the same time, I find an interesting comment by Ebbinghaus. He had been goaded by Dilthey, who had mounted an attack on *erklärende Psychologie*.[2] Dilthey objected to the *atomistische Zersplitterungen* which he felt was the mark of the New Psychology. To this Ebbinghaus replied (my translation) that the psychologist "speaks of sensations, perceptions, feelings, associations, as if they were simple and complex constructions and functions. But he does not assert thereby that the complex construction is derived from the simple, or that the whole is created by putting together complicated parts."[3] There is more in the same vein. Ebbinghaus then draws a distinction:

> It is indeed not true that the mind because of its unity is to be considered as something different from external objects. The contrast exists only with stones and clouds, in general with the things with which physics and chemistry are principally concerned. The contrast is not at all with the organic world, with the plants, with animals, and the human body. For these, all that was said about the unity of the mind is equally true.[4]

He makes this point more explicit:

> It is certainly true that the first form of modern scientific psychology, the so-called association psychology, does not come up to a much to be desired standard of perfection, but suffers rather from evident deficiencies. These go

[1] These formed part of the Langfeld Lectures given by Köhler at Princeton University in the fall of 1966 and since published as W. Köhler, *The Task of Gestalt Psychology*, 1969.

[2] W. Dilthey, Ideen über eine beschreibende und zergliedernde Psychologie, *Sitzber. Berl. Akad. Wiss.*, 20 December 1894, 1308–1407.

[3] H. Ebbinghaus, *Grundzüge der Psychologie*, 1897, 166.

[4] Ebbinghaus, *op. cit.*, 165.

back to two roots: Psychology has put its trust too greatly in the theoretical mastery of psychological facts, and it has followed to too great an extent physicochemical analogies. Both are understandable. . . . And where else could it [psychology] find a concrete example of the procedures of true and productive science than in physics and chemistry, since a significant biology had not yet been developed? Nevertheless, understandable or not, the deficiency exists. It consists, not exclusively, but essentially, in the unsatisfactory evaluation of the peculiar unity or whole, if the expression be permitted, into which the differentiated many are assembled and combined in mental life.[5]

Ebbinghaus' comment poses a very interesting question. It was precisely to physics and chemistry, especially physics, to which the Gestalt psychologists were appealing in support of their views about whole properties in psychology. Was this indeed the same physics which, according to Ebbinghaus, offered the false analogies? More particularly, what was the physics and chemistry, to which Ebbinghaus had reference, that had led psychology astray? The answer to this question is not easily discovered. If one reads the systematic texts of this period, there is practically no clue. About the biological side there is no question. In Lotze, in Spencer, in Wundt, in Ebbinghaus, in James, there is the long and deep obeisance before the Nervous System.[6] The graduate student of that day could be depicted with a color wheel in one hand and a sheep's brain in the other. But the treatment of physics is completely incidental and undocumented. Where did Wundt or Külpe or James or Brentano learn about the *echte und fruchtbare Wissenschaft?*

One important point of contact was, of course, Helmholtz. He collaborated in founding the *Zeitschrift für Psychologie* and appeared as an author five times in the seven volumes published before his death. The last paper he wrote was published in the *Zeitschrift*.[7] But it is hard to escape the feeling that physics itself was passing him by. Helmholtz was obviously a key figure in color theory and in hearing. It is less evident that physical theory as such was brought to bear on psychological problems. At two points it was. The notion of conservation of energy had a profound effect in sharpening the debate about the mind-body problem.[8] To this

[5] H. Ebbinghaus, Ueber erklärende und beschreibende Psychologie, *Z. Psychol.*, 9, 1895, 175. This entire article (161–205) is a recapitulation and reply to Dilthey's attack.

[6] I have reference to the space accorded to neuroanatomy and related topics in the general textbooks of these authors. Detailed documentation is scarcely required.

[7] H. von Helmholtz, Ueber den Ursprung der richtigen Deutung unserer Sinneseindrücke, *Z. Psychol.*, 7, 1894, 81–96.

[8] H. von Helmholtz, *Ueber die Erhaltung der Kraft: Eine physikalische*

issue Helmholtz had made an early and telling contribution. The second was Helmholtz' elucidation of Riemann space, which he used hopefully to support an empiristic theory of space.[9] Neither of these spoke particularly to the issue raised by Ebbinghaus.

It seems fair to conclude that the model of the physical world which nineteenth-century psychology, at least nineteenth-century German psychology, sought to copy was, in its principal outline, Newtonian. This is an inference for which I have discovered no direct evidence, but I believe it is a plausible inference. I should like to elaborate my reasons for believing this.

First, what are the elements of Newton's mechanics that are important in this context? We remember that Newton accounted for the behavior of the moon with his notion that the moon was held in its orbit by the force of gravity. It remained to show that the force varied with the inverse square of the distance and directly with the product of the two masses that were attracting one another. To do this, Newton argued: (1) that there is an absolute space described by Euclidean geometry and infinite in extent; (2) that this space is in general a void except for numberless discrete objects or particles moving or resting within it; (3) that the objects in the space are subjected to certain forces, in part, of an unknown origin; (4) that one of these forces exhibits itself as gravity which is a mutual attraction of two objects with a force proportional to $M_1 M_2 / d^2$; (5) that the forces of gravity between pairs of objects are independent of one another and that the motion of a body is the result of all the forces acting on it in combination at any one point in time. There is, in addition, the concept of inertia and implicitly the notion of conservation of matter. These are the key elements: the fixed but empty frame of space, particulate matter moving in that space, forces acting at a distance, and the interaction of these forces by arithmetic combination.

Much of the Newtonian scheme was, of course, ready at hand in the rapidly growing knowledge of astronomy. Newton ordered

Abhandlung, 1847. Reprinted substantially in *Fortschritte der Physik, im Jahre 1847*, 3, 1850, 233–245. Expanded in three lectures in *Medical Times and Gazette* (London), 1864, and in *Populäre wissenschaftliche Vorträge*, 2–3, 1876. See the many references given by Rand in J. M. Baldwin, *Dictionary of Philosophy and Psychology*, 3, 1905, 1091–1099. For the comment of a physicist, see E. Mach, *The Analysis of Sensation*, S. Waterlow (trans. *Analyse der Empfindungen*, 5th ed., 1906), 1959, 55.

[9] H. von Helmholtz, *Ueber den Ursprung und die Bedeutung der geometrischen Axiome*, in *Populäre wissenschaftliche Vorträge*, 1870. Also in *Vortäge und Reden*, 2, 1884.

these observations with great skill. But perhaps more fundamentally, he broke with the tradition that sought a substantial agent to explain any action, the important vestige of primitive magic. Gravity, acting through empty space, took the place of the invisible shell holding the planets in place.

Newton's theories had become a commonplace long before psychology was born. It was certainly the dominant view by the middle of the eighteenth century. But by the beginning of the nineteenth century, there were cracks in the edifice. There had from the beginning been Huygens' wave theory of light. Then came Faraday, Bolyai, and Lobatschevsky.[10] Try as men would, they could never put Humpty Dumpty together again. Then why did psychology follow Newton—psychology that was taking form a half century after the Newtonian structure started to crumble? The evidence is that it did. It accepted an a priori space. It created a set of particulate and independent elements. It made use of arithmetic rules of combination. Why? The reason, I believe, is because of one man, Immanuel Kant. Both the anachronistic physical theory and the absence of any clear account of that *echte und fruchtbare Wissenschaft* is a result of the fact that psychologists took their physics not from physicists but from a philosopher.

There is relatively direct evidence for the connection. Scholars of Kant have made explicit reference to Kant's admiration for Newton and for his acceptance of the Newtonian view of the world.[11] In fact, only the supposition that Kant took Newton very seriously will explain the peculiar position afforded to judgments about space in Kant's system: space is never the object of an analytical judgment; it is a pure intuition, the only pure intuition. The propositions of geometry are propositions about space, and they are absolutely true; therefore, judgments about space are a priori judgments, or synthetic judgments.

The second step is to note that almost every psychologist who wrote about perception, or even psychophysics, wrote with one eye on what Kant had said. In nineteenth-century psychology, Kant is as ubiquitous as Pavlov and Freud in the twentieth century. Thus, the doctrine of local signs, from Johannes Müller to Külpe and

[10] For a brief but persuasive account of the basic character of this change, see A. Einstein, Maxwell's influence on the development of physical reality, in J. J. Thomson *et al.*, *James Clark Maxwell: A Commemoration Volume, 1831–1931*, 1931.

[11] See N. K. Smith, *A Commentary to Kant's "Critique of Pure Reason,"* 1923, lv–lvi. I have, in general, followed Smith.

Titchener, is a direct consequence of Kant's dictum: "Space does not represent any quality of objects by themselves, or objects in their relation to one another."[12] One hundred years of speculation and relatively fruitless experimentation rest upon this single point. It is evident, of course, that the detailed knowledge of physical phenomena—of optics, of acoustics, and of electricity—did not come from this source. The problem lies precisely in the fact that theory and practice came from such different sources. Psychologists used simple physical facts without knowing anything of Faraday and Maxwell, of Clausius and Kirchoff, of Rutherford and Lorentz.

I have to confess that there are gaps in my view. There are aspects of Newtonian theory, particularly the concept of force, for which there is no easy equivalent in the Kantian view. In this area, and with regard to the concept of energy, a strong case can be made for quite other lines of influence. How else could psychologists have conscientiously defended a theory of psychophysical interaction in the face of the notion of conservation of energy? Or again, there was Helmholtz volubly expounding models of non-Euclidean space without there being any very constructive response by psychologists to the problems this raised. The empiristic theory of space on which some psychologists came to depend was really more of a nontheory than a successful, positive account.

In conclusion, let me return to the elements of Newtonian theory. This was the model to be followed. At one point, Einstein characterized them this way: "According to Newton's system, Physical Reality is characterized by the concepts space, time, the material particle, and force. . . ."[13] Space for Newton was an absolute space, empty; a given, existing independently of the events that occurred within it. For Kant this became an a priori space, a pure intuition not based on experience, preexisting, and serving as the framework within which objects might be related to one another. For the nineteenth-century psychologist, space was a construct, not something given directly in experience, but rather a network of nonspatial local signs knitted together by nature *or* by associations.

Let me take another aspect. For Newton there were bits of matter, particles, moving about in the void of space. For Kant there were objects, the *Dinge an sich* which gave rise to sensory impressions, qualitative in nature. By an analytic judgment there was formed a concept of the object, still basically qualitative in nature. The psychologist took the view that there were clusters

[12] I. Kant, *Critique of Pure Reason*, F. M. Müller (trans.), 1881, A26, B42.

of sensations to which might be added by association other sensorial elements to form larger complexes.

Time for Newton was independent of space although it would be measured by motion and, accordingly, assuming constant velocity, by measures of space. Time for Kant was subjective and internal. It depended upon judgments of succession and simultaneity. It was relational and was therefore a priori in nature. It was, however, unlike space, something derived from the manifold of sensations and therefore not a pure intuition, as was space. Time for the psychologist was much the same thing. It depended upon the flow of experience, upon judgments of succession and simultaneity. It became represented by qualitatively heterogeneous experiences through processes, again, of association.

The answer, then, to the Ebbinghaus question as to why psychology had no way of dealing with the "unity of the mind" does not lie, in my opinion, in the fact that psychology took physics and chemistry as its model. Each of the elements in the Newtonian scheme had come under strong attack by the end of the nineteenth century, and a very different conceptual scheme was emerging, so far as physics was concerned. The problem was not so much that psychology was wrong in taking physics as a model. The problem was that it took the wrong physics as a model.

One last comment. The point that the Gestalt psychologists were making was not, in my opinion, that psychology should reject physical models in favor of a loose phenomenology or the mythology of psychoanalysis. Rather, psychology should take better and more meaningful physical models. This point has been overlooked and neglected. But that is another discussion.

Harvard University

EDWIN B. NEWMAN

A FOOD DISH FOR TESTING FOOD AND STIMULATION REWARD WITH SMALL ANIMALS

Animals will bar-press for brain stimulation as reward.[1] It is desirable to compare certain aspects of this bar-pressing behavior,

* Received for publication February 18, 1968. The research reported was supported by grants to R. B. Malmo from the following sources: National and Medical Research Councils and Defence Research Board of Canada; and U.S. National Institute of Mental Health and National Science Foundation.

[1] J. Olds and P. Milner, Positive reinforcement produced by electrical stimula-

and its physiological accompaniments, with corresponding phenomena associated with bar-pressing for nutritive reward (*e.g.* food or water). In order to obtain this comparison in the behavior of a rat, a food dish has been developed which allows the rat to (*a*) obtain a food pellet after a bar-press or (*b*) to receive electric brain stimulation after a bar-press, but only when its head reaches into the dish. Thus, for both the food reward and stimulation reward, the animal must execute similar motions.[2] Tested were food-deprived rats; their behavioral responses were assessed by the number of bar-presses and changes in heart rate.[3]

The food dish is fitted with metal electrodes which form a capacitor, and the principle of operation is such that by putting its head in the dish, the animal changes the permittivity (dielectric constant) of the capacitance and this change is detected by an A.C. bridge network. The output signal of the bridge, after electronic amplifying and shaping, is applied to equipment for behavioral programming.

Circuit

Fig. 1 shows the circuit diagram. A Hartley-type oscillator supplies the excitation voltage to the bridge network, a 'Blumlein bridge.' This type of bridge is particularly well suited for measuring small capacitances because of its sensitivity and zero stability, these characteristics being achieved by virtue of its inductively coupled ratio arms.[4] The use of miniature transformers (T_1 and T_2) allows the bridge network and bridge amplifier to be built into a small metal enclosure and mounted on the outside of the test box, near the food dish. When the food dish is empty, the bridge is balanced by means of a variable capacitor (C) for a minimum output signal. The output voltage of the bridge increases with any variation of the capacitance of the food dish. The signal is monitored with an oscilloscope or sensitive voltmeter connected to the Signal Monitor output at Emitter Follower 1.

The amplifier consists of two common emitter stages and its performance is stabilized by two feedback loops. A current-shunt feedback from emitter of Q_2 via transformer winding to the base

---

tion of septal area and other regions of rat brain. *J. comp. physiol. Psychol.*, 47, 1954, 419–427.

[2] This experiment was designed by G. Blevings.

[3] W. J. Mundl, A cardiotachograph with logarithmic response for use with rats, *Psychophysiology*, 3, 1966, 220–223.

[4] A. F. Giles, *Electronic Sensing Devices*, 1966.

of Q_1 provides D.C. stabilization and biasing of Q_1. Voltage-series feedback between collector of Q_2 and emitter of Q_1 provides for both A.C. and D.C. stabilization. This latter feedback seemingly contradicts the notion of negative feedback, since the signal fed back is actually in phase with the signal at the emitter of Q_1. Nevertheless, its effect is that of negative feedback since it provides opposition to the input signal by decreasing base-to-emitter bias in phase with the input signal. Or, alternatively, Q_1 may be looked upon as a common base stage in the feedback loop, the only phase reversal taking place in Q_2.

The Zener diode eliminates a coupling capacitor between the stages. Depending on the trigger adjustment, the Schmitt trigger becomes activated at a certain signal level; this trigger adjustment therefore controls the sensitivity of the instrument. The output pulses of the Schmitt trigger are integrated by RC networks before and after Emitter Follower 2 to provide a voltage envelope which is applied to Q_3 to drive it into saturation. Outputs in the form of a contact closure or pulse are available for activating programming equipment.

Food Dish

The inside part of the food dish was cut from a plastic bottle of oval shape (see Fig. 2). A ring of thin brass sheet was fitted around

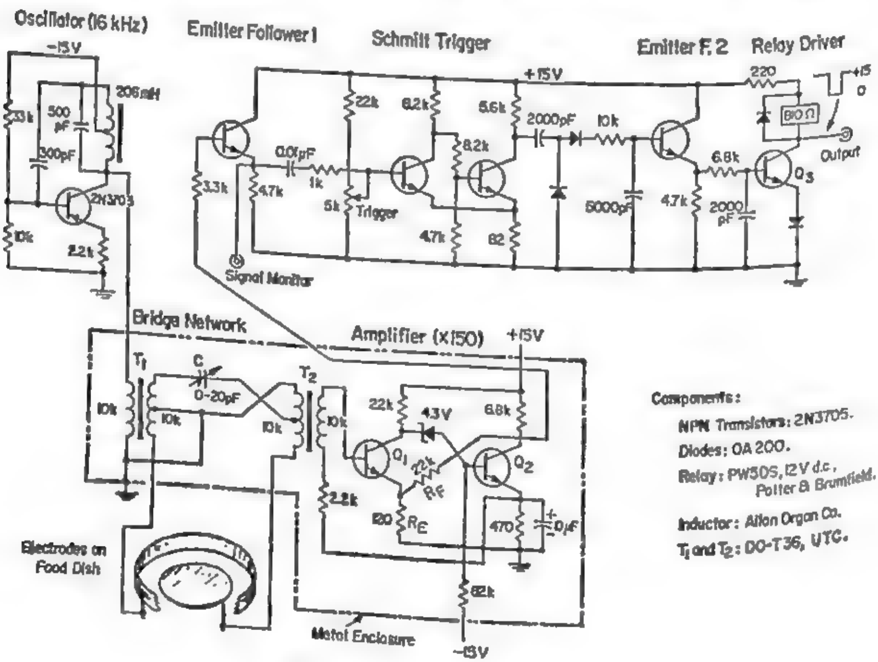


Fig. 1. Circuit Diagram

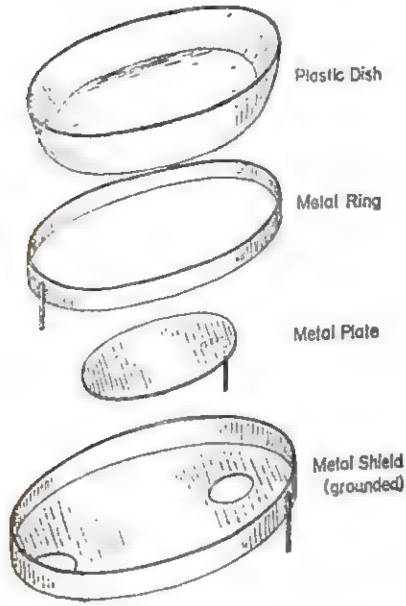


FIG. 2. EXPLODED VIEW OF FOOD DISH

the plastic dish and represents one electrode of the capacitor. The other electrode consists of a brass plate cemented to the underside. A larger metal dish encloses the bottom half of the dish-electrode assembly and serves as ground shield. This metal dish should be large enough to allow for an approximately ¼-in. space between it and the electrodes. The ground shield eliminates unwanted operation of the instrument in situations where the animal could initiate triggering from the outside of the dish by its proximity to the electrodes. The assembled parts are held in place with silicon rubber, which is also applied freely to the outer surface for electrical insulation and physical protection.[5]

PERFORMANCE

Operation of the instrument was found to be stable. No adjustments were needed once the bridge circuit was balanced and the trigger adjusted for the required sensitivity. A difference in performance must be expected between an animal wired to recording or stimulation equipment and an electrically isolated animal. A connected animal is by necessity at ground potential—through resistive or capacitive coupling at the instrument inputs—and this factor contributes additional imbalance to the bridge network. Some addi-

[5] Silastic 732 RTV, Dow-Corning.

tional sensitivity can be obtained by decreasing feedback in the amplifier, *i.e.* by raising the ratio of RF/RE.

Allan Memorial Institute,
McGill University

WILLIAM J. MUNDL

APPARATUS FOR RANDOMLY DETERMINING VISUAL STIMULUS POSITIONS USING A 'FORCED-CHOICE' PSYCHOPHYSICAL PROCEDURE

Several studies of visual perception involving tachistoscopic presentations have been carried out in this laboratory using a simple 'forced-choice' psychophysical procedure.[1] The subjects were required to discriminate the spatial orientation of a figure, such as a semicircle, illuminated by brief flashes of light. This procedure offers two principal advantages over other forced-choice techniques, which have required *S* to discriminate different stimulus forms[2] or to discriminate from among several different stimulus presentations.[3] First, stimulus presentation is simplified, because the same stimulus figure is always used, and only its orientation need be changed. Second, the use of a single stimulus figure eliminates problems concerning the differential detectability of stimulus options (such as can occur using letters of the alphabet). The apparatus described here was designed to rotate stimulus

* Received for publication February 3, 1969. The research reported was supported in part by NASA Grant NGL 05-007-049 (formerly NsG-623) and NSF Grant GB-1844.

[1] R. B. Boyle, An investigation of the latency hypothesis of perceptual interference resulting from successive visual presentations, unpublished doctoral dissertation, University of California, Los Angeles, 1963; M. L. Kietzman, The perceptual interference of successively presented visual stimuli, unpublished doctoral dissertation. University of California, Los Angeles, 1962; Emanuel Donchin, J. D. Wicke, and D. B. Lindsley, Cortical evoked potentials and the perception of paired flashes, *Science*, 141, 1963, 1285–1286; Emanuel Donchin and D. B. Lindsley, Retroactive brightness enhancement with brief paired flashes of light, *Vision Res.*, 5, 1965, 59–70; Emanuel Donchin and D. B. Lindsley, Visual evoked response correlates of perceptual masking and enhancement, *Electroenceph. clin. Neurophysiol.*, 19, 1965, 325–335; Emanuel Donchin, Retroactive visual masking: Effects of test flash duration on the masking interval, *Vision Res.*, 7, 1967, 79–87; Emanuel Donchin, Average evoked potentials and uncertainty resolution, *Psychon. Sci.*, 12, 1968, 103.

[2] C. W. Eriksen and M. Hoffman, Form recognition at brief durations as a function of adapting field and interval between stimulations, *J. exp. Psychol.*, 66, 1963, 485–499.

[3] H. R. Blackwell, Psychophysical thresholds: Experimental studies of methods of measurement, *Univ. Mich. Eng. Res. Bull.*, 1953 (No. 36), 1–227.

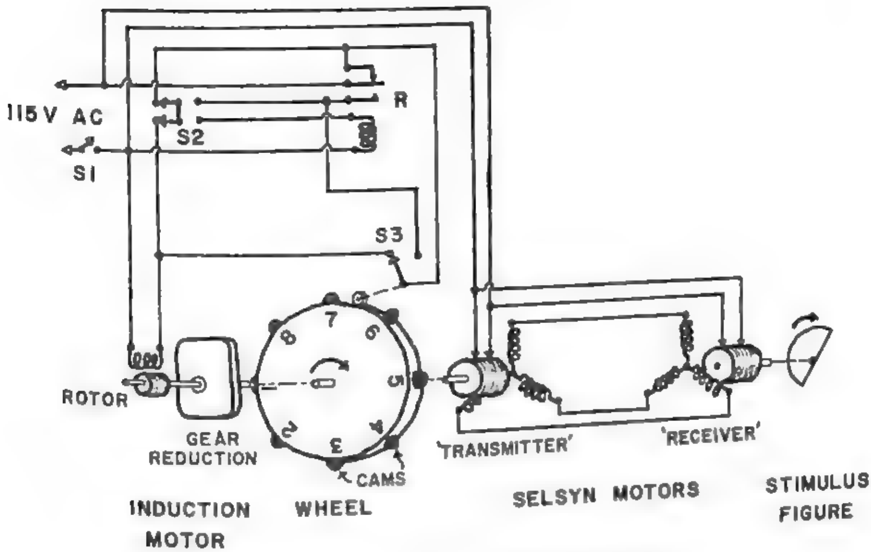


FIG. 1. SCHEMATIC DIAGRAM OF APPARATUS

figures randomly into any of eight different positions, separated by 45°.

Fig. 1 shows a schematic diagram of the system. On the left side is the control unit which determines the orientation of the stimulus figure on a given trial. A geared-down induction motor[4] turns a wheel on which cams have been placed for each of the eight different stimulus positions. With the power turned on (S1), the induction motor turns continuously until *E* presses the foot switch (S2). This allows the roller-actuated Micro switch (S3) to provide a brief pulse of current at the next upcoming cam which is sufficient to activate the holding relay (R). Through the action of R, power is removed from the induction motor, as well as being applied to the field coil of R itself, in parallel with that from S3. This prevents the induction motor from receiving power again before S2 is released. Once its power is turned off, the induction motor is quickly stopped by the action of a magnetic brake (not shown in the figure). The magnetic brake is a integral part of the motor, and it is unique in that it requires no additional power because it is operated by the magnetic field of the motor coil itself. The stimulus position at which the control unit has stopped may be identified by noting

[4] About 60–80 rpm, with continuous-duty field coil, and built-in magnetic brake: Speedway Mfg. Co., La Grange Park, Ill., or Breval Products Corp., Carlstadt, N. J.

the identification number or letter marked on the wheel for the cam that actuated S3.

The remainder of the system consists of two selsyn motors (Type VII), which remain on at all times by the action of S1. One of these, the 'transmitter' selsyn, is mechanically connected to the drive shaft of the induction motor so that it rotates and stops with the wheel and cams. This selsyn is electrically connected to the second, or 'receiver,' selsyn motor, whose drive shaft is connected to the stimulus figure. Any movement initiated by the control unit is picked up by the 'transmitter' selsyn and is faithfully reproduced by the 'receiver' selsyn.

Random selection of the stimulus positions is determined as follows: The induction motor is allowed to run for several seconds, and then *E* stops it by pressing S2. The motor remains stopped until *S* is given a chance to view the stimulus (*e.g.* by means of a flash of light which occurs when he presses a switch connected to other circuitry) and indicates the stimulus orientation. *E* records both the *S*'s response and the cam identification shown on the wheel and then releases S2, which allows the motor to run again. After several seconds—usually the time it takes to score the response—*E* presses S2 again and starts a new trial. The induction motor rotates rapidly through the eight possible stimulus positions, and a new position occurs about once every $\frac{1}{10}$ sec. The variability in the time it takes to record *S*'s response is sufficient to ensure that the stimulus positions will be determined randomly. This has been verified numerous times using many different subject-experimenter combinations.

This system was designed to rotate small field stops which formed stimulus figures when inserted into a Maxwellian-view optical system. The field stop was mounted in ball bearings, and rotated by means of a 1:1 gear train attached to the 'receiver' selsyn motor. It can, however, find application in almost any visual-perception apparatus where stimuli are illuminated for brief periods. Only the linkage between the 'receiver' selsyn and the stimulus figure need be modified to suit individual requirements. The accuracy with which a stimulus figure may be rotated to a given position is about ±2° because of variation in the time it takes for the induction motor to be stopped. While this error is small, relative to the 45° spacing between adjacent stimulus positions, and did not affect the discrimination of stimulus positions, it could be reduced through

the use of a more sophisticated, and considerably more expensive, clutch-brake assembly.

University of California, Los Angeles

J. D. Wicke
D. B. Lindsley

THE SIXTY-FIRST ANNUAL MEETING OF THE SOUTHERN SOCIETY FOR PHILOSOPHY AND PSYCHOLOGY

The sixty-first annual meeting of the Southern Society for Philosophy and Psychology was held on April 3–5, 1969, at the Everglades Hotel in Miami, Florida. Hosts were the Departments of Philosophy and Psychology at the University of Miami. Approximately 325 members, friends, and guests were in attendance. The program consisted of two symposia in psychology and two joint sessions. All together, there were 32 papers in philosophy and 57 papers in psychology. As usual, a social hour and business meeting were also included in the program.

At the first of the two joint sessions, Earl A. Alluisi, University of Louisville, introduced E. M. Adams, University of North Carolina, who delivered the presidential address, "The Scientific and the Humanistic Images of Man in the World." At the second joint session, Paul L. Adams, University of Florida, delivered the invited address, "Professors and Citizen Activism." Discussants were Thomas K. Hearn, Jr., College of William and Mary, representing philosophy, and Denis O'Donovan, Florida Atlantic University, representing psychology.

Four Junior Awards were given. Corecipients in philosophy were James F. Harris, Jr., University of Georgia, for his paper "Quine on Analyticity and Logical Truth," and John Beversluis, Emory University, for his paper "'I Know': An Illocutionary Analysis." Corecipients in psychology were Virgil V. McKenna, College of William and Mary, for his paper "Stylistic Factors in Learning and Retention," and Ben B. Morgan, Jr., University of Louisville, for his paper "The Acquisition of a Problem-solving Skill." These awards are given annually for the papers of highest merit presented by members who have not yet received the doctoral degree or who have received it within the last five years.

Cochairmen of the Committee on Local Arrangements were Gerrit

Schipper, University of Miami, for philosophy, and Daniel Lordahl, University of Miami, for psychology. Cochairmen of the Program Committee were James W. Oliver, University of South Carolina (philosophy), and William Dawson, University of Florida (psychology). Cochairman of the Junior Awards Committee were Willis Moore, Southern Illinois University (philosophy), and Earl A. Alluisi, University of Louisville (psychology). At the annual business meeting, reports of the officers and of the several standing and ad hoc committees were received and accepted. It was announced that the Council had accepted the invitation of the Department of Psychology at Washington University to hold its sixty-fourth annual meeting in Saint Louis, Missouri, March 30–April 1, 1972.

The following officers were elected to serve during 1969–1970. Malcolm D. Arnoult, Texas Christian University, President; Dwight Van de Vate, Jr., University of Tennessee, President-Elect; E. M. Adams, University of North Carolina, Past President; William T. Blackstone, University of Georgia, Secretary. Andrew J. Reck, Tulane University, continues to serve as Treasurer. Douglas Browning, University of Miami, and Donald S. Lee, Tulane University, were elected to the Council for three-year terms as representatives of philosophy. L. J. Peacock, University of Georgia, was elected to the Council for a three-year term as representative of psychology. Other continuing members of the Council are John W. Davis, Milton H. Hodge, and Michel Loeb.

The sixty-second annual meeting of the Southern Society for Philosophy and Psychology will be held on March 26–28, 1970, in Durham, North Carolina. It will be hosted by the Department of Philosophy of the University of North Carolina.

University of Miami DOUGLAS BROWNING

THE AMERICAN
JOURNAL OF PSYCHOLOGY

Founded in 1887 by G. STANLEY HALL

Vol. 82 DECEMBER 1969 No. 4

THE ROLE OF THEMATIC RELEVANCE OF CONTEXT IN PAIRED-ASSOCIATES LEARNING

By WILLIAM BEVAN and STANLEY C. COLLYER,
Johns Hopkins University

In a 1964 paper in this journal, Postman and Phillips reported that when context words accompanied a stimulus word in a paired-associates task, learning was retarded to the extent that these background stimuli were meaningful.[1] Subjects receiving context words with a high frequency of usage required, on the average, two trials longer to reach a criterion of 8/10 correct than did subjects given low-frequency background stimuli. The authors hypothesized that the greater the meaningfulness of the contextual stimuli, the greater the likelihood of associations that compete with those prescribed in the definition of the task. It should be noted, however, that the general effect of context upon performance is unclear, since Postman and Phillips did not provide information on performance under a context-free control condition.

The present experiment, which included control groups, asked if differences in thematic relevance among meaningful contextual stimuli are related to their effectiveness in influencing acquisition, and compared the influence of contextual stimuli when they are related to the focal stimulus word, on the one hand, and to the specified response word, on the other. A consistent thematic relationship between contextual and focal stimulus words provides a basis for mutually reinforcing multiple associations with the response word, while

* Received for publication April 1, 1969. The experiment reported was performed under Contract N00014-67-A-0163-0001 between the Physiological Psychology Branch, Office of Naval Research, and The Johns Hopkins University.

[1] Leo Postman and Laura W. Phillips, The effects of variable context on the acquisition and retention of paired associates, this JOURNAL, 77, 1964, 64–74.

a consistent thematic relationship between context and response words should constitute a general cue to the identity of the latter. The reader will undoubtedly find this study reminiscent of Pan's classical study of context effects in learning,[2] with procedural characteristics similar to those of Postman and Phillips, except that in the present experiment a particular context word was associated with only one stimulus-response pair. The purpose of the present study may be briefly summarized by these questions: In general, does context enhance or degrade learning? How does the 'cohesiveness' of the stimulus-context unit, as defined by the size of the smallest class of elements to which both words belong, affect the learning of a response to a member of that class? Is the ability of a word to serve as a contextual cue to a response related to later performance with the cue removed? When context words are present only as rote stimuli, what effect do they have on learning, and what implications does this have for the role of meaningful context in learning?

Method

Experimental design. Ten independent groups of *S*s were employed. These are identified in Table I, which shows how the groups were divided into two major categories, according to whether contextual words stood in relation to the stimulus item (C_S groups) or to the response item (C_R groups) of the S-R pair. Within each category four types of context were employed, along with one control group without context stimuli. In the *High* group the level of relationship between context words and the stimulus or response word was maximal (*e.g.* all context words were names of types of footwear, of flowers, of alcoholic beverages, of automotive vehicles, of meat, of houses of worship, of card games, feline animals, etc.). In the *Medium* group the contextual items were selected to bear a meaningful, but more general, relation to the stimulus or response word (*e.g.* context words were types of articles of clothing, of plants, of beverages, of modes of transportation, food, buildings, games, animals, etc.). The *Random* group consisted of context words carefully selected to bear no logical and consistent relationship to the stimulus or response word with which they were paired. A fourth context group involved the use of a single word, 'nutshell,' as the context word in all instances. This was included in order to determine the role of contextual items with the effects of thematic relevance removed. This group was labeled the *Invariant* group. Finally, a *Control* group served to set a performance baseline in the absence of context words.

It will be noted that this procedure is similar to that employed by researchers in the area of stimulus selection,[3] where interest is in the extent to which people abstract or condense stimulus information from a complex array. The present technique, however, would be an atypical stimulus-selection paradigm

[2] Shuh Pan, The influence of context upon learning and recall, *J. exp. Psychol.*, 9, 1926, 468–491.

[3] See, for example, B. J. Underwood's chapter on rote learning in A. W. Melton (ed.), *Categories of Human Learning*, 1964, 47–78.

TABLE I

EXPERIMENTAL DESIGN

Condition	Level of relationship	Example S	Example R
Context related to stimulus (C_S)	High (C_SH)	TIGER LION	GIH
	Medium (C_SM)	ELEPHANT LION	GIH
	Random (C_SR)	HAMMER LION	GIH
	Invariant (C_SI)	——† LION	GIH
	Control (C_SC)	LION	GIH
Context related to response (C_R)	High (C_RH)	TIGER GIH	LION
	Medium (C_RM)	ELEPHANT GIH	LION
	Random (C_RR)	HAMMER GIH	LION
	Invariant (C_RI)	——† GIH	LION
	Control (C_RC)	GIH	LION

† Pronounce 'nutshell.'

due to the fact that the compound stimulus was composed of only one constantly recurring item, along with several variable ones.

Subjects. A total of 134 Johns Hopkins undergraduates and 10 from the University of Maryland participated as paid volunteers in the study. Twenty-four of these were discarded, consistent with the procedure described below.

Materials. The paired-associates material consisted of 15 English nouns (median frequency in the Thorndike-Lorge General Count was 7; range 1-AA) and 15 CVCs of low association value (range 1%–24% in the Archer List).[4] For each of the 15 nouns, four context nouns were selected at each of the three levels of thematic relationship described above (*e.g.* LION—TIGER, PANTHER, COUGAR, LEOPARD; LION—ELEPHANT, ZEBRA, GIRAFFE, GORILLA; LION—HAMMER, MIMEOGRAPH, BUCKLE, CLOCK). The 180 context words were roughly matched across levels of frequency of usage, except for a few not included in the Thorndike-Lorge list.

All material was typed in black and mounted on white plastic cards for presentation in a Hunter Card-Master, a device which performs the same functions as a memory drum. For each context condition (except Invariant), four decks of 15 cards were prepared, each deck containing the 15 S-R pairs, but with 15 different context items. For a particular S-R pair across the four decks, context appeared above the stimulus on two decks, and below on two, while the response was located directly opposite the stimulus term. Within a deck roughly half the context items were above the stimulus and half below.

[4] E. L. Thorndike and I. Lorge, *The Teacher's Wordbook of 30,000 Words*, 1944; E. J. Archer, A re-evaluation of the meaningfulness of all possible CVC trigrams, *Psychol. Monogr.*, 74, 1960 (Whole No. 497).

Each trial involved the use of one deck, and decks were used in sequence, so that the first deck was used on Trials 1, 5, 9, etc. Each deck was reordered every time it was used. In the Invariant groups the context consisted of a heavy black line above or below the stimulus, to which the S was instructed to respond with the word 'nutshell.' Only two decks of cards were required for these groups.

Procedure. The method of testing was that of anticipation. For all but the two Control groups, each combination of focal stimulus and context was exposed on one side of the viewing apparatus for 5 sec. followed by a 2-sec. interval in which the response was exposed as well. Between each paired-associates sequence, 3 sec. intervened. S was requested to read from top to bottom, pronouncing words and spelling out CVCs. If S was unable to anticipate the response within 5 sec., he gave it when it appeared. Procedure for the Control groups was the same as for the other groups, except that the anticipation interval was reduced from 5 to 4 sec., to compensate for the fact that no contextual item was pronounced.

A number of Ss failed to reach the criterion of one errorless trial (15/15 correct) within the 1-hr. experimental session, and they were disqualified. They were replaced by new Ss by means of a substitution technique developed by B. J. Underwood.[5] The largest number of nonlearners in any group of 10 Ss was four; these Ss were replaced and the N of all other groups (except C_R High and Medium) was inflated to 14, and then reduced to 10 by excluding the four slowest learners, including all nonlearners. The resultant 10 Ss per group were thus not representative of the entire population from which they were drawn but rather of only the upper 71% of that population. It was deemed unnecessary to add Ss to the C_R High and Medium groups since learning rates for these two groups were sufficiently greater than those for the other groups that such a refinement was not necessary to show differences between groups.

Immediately following attainment of one perfect trial, Ss were retrained to the same criterion of 15/15 with context absent. This change in procedure was introduced without prior warning.

Results and Discussion

Performance in the Control groups. Trials to criterion during initial learning for all groups and during relearning for the context groups are presented in Table II. A comparison of the two Control groups provides information about the relative difficulty of the task under the C_S and C_R paradigms. It is seen that under either condition the average S achieved errorless anticipation of all response items in less than 10 trials or after about 25 min. of training. Furthermore, it was not reliably easier for S to respond to a noun with a nonsense syllable than to a noun with a nonsense syllable [t (18) = .19, $p > .05$]. This is at variance with the common expectation that associations are more readily formed when the initial member of the

[5] B. J. Underwood, personal communication.

TABLE II

Trials to Criterion (15/15) as a Function of Level of Thematic Relation between Context Words and either Stimulus or Response

Condition	Level of relationship	Initial learning with context	Relearning without context
Context related to stimulus (C_S)			
C_SH	High	10.1	1.9
C_SM	Medium	10.2	1.3
C_SR	Random	11.9	1.4
C_SI	Invariant	10.4	
C_SC	Control	8.6	
Context related to response (C_R)			
C_RH	High	3.2	9.0
C_RM	Medium	6.0	8.1
C_RR	Random	9.9	1.9
C_RI	Invariant	10.1	
C_RC	Control	8.4	

common-word–nonsense-syllable pair is the nonsense syllable rather than the word. It will be recalled that these data were obtained after the four *Ss* per group had been discarded. Since, however, there were no nonlearners in the original sample for either of the Control groups, it is not necessary to discard data in order to make simple comparisons between these groups. Furthermore, the lack of difference between the groups prompted the testing of additional *Ss* under these conditions to evaluate the validity of these results. With the *N* per Control group raised to 18 (20 with 2 per group discarded to compensate for nonlearners in the subsequent sample), performance with the nonsense-syllable–common-word sequence was better than with the reverse order by 1.3 trials; this difference, however, was still not significant [t (34) = 1.40, $p > .05$].

We can only speculate as to why this study failed to find a significant difference in performance under these two conditions. Several possibilities come to mind. First, the requirement that *S* spell all CVCs aloud may lessen the operation of stimulus selection as a helpful strategy. Pronouncing all three letters as a unit could conceivably make it more difficult to employ the frequently preferred strategy of focusing on a single letter (or at most, two) as the functional stimulus.[6] Second, high interstimulus similarity increases task difficulty since a single letter is insufficient to discriminate between

[6] B. J. Underwood, Stimulus selection in verbal learning, in C. N. Cofer and Barbara S. Musgrave (eds.), *Verbal Behavior and Learning*, 1963, 35.

stimuli. In this study similarity among CVCs was moderately high, with four pairs sharing the same first letter, two triplets and a pair with the same final letter, and up to four stimuli with the same middle vowel. Similarity among nouns, however, was low, as the nature of the selection procedure required that they be quite different in meaning. Third, it is possible that slight differences in the effective anticipation interval resulted from the likelihood that it takes slightly longer to pronounce three separate letters than it does to say a two-syllable word. This could mean less recall time available in the C_R condition, hence a reduced learning rate. It is difficult to know how much importance to attach to such a factor, but it is possible that it could account for these results.

Whereas no differences were found between the groups on initial learning, there was a highly significant difference on a single reversal trial immediately following the attainment of criterion. Here *S*s were shown the response terms at the same presentation intervals and requested to recall the corresponding stimulus terms. The average number of items correctly recalled for the C_S group, which now received the CVC as a stimulus, was 13.6, while that for the C_R was 10.8 [t (34) $= 3.94$, $p < .001$]. While no careful record was kept of the several types of errors which occurred, it was observed that C_R *S*s frequently erred on only one of the three letters. Thus, it would appear that *S*s originally learning CVC-noun combinations were not in fact responding to the stimulus term as a whole but rather to some portion of it.

Performance with context and stimulus words thematically related (the C_S groups). Table II indicates that when the thematic relationship between context and stimulus words was minimal (*i.e.* when the context word had been randomly paired with the stimulus), performance was degraded about 40% [C_SR vs. C_SC: t (18) $= 2.97$, $p < .01$]. However, when the context and stimulus words were thematically related, the interference associated with the presence of the context word was attenuated. In the High and Medium groups, performance was degraded around 20% [C_SH vs. C_SC: t (18) $= 1.12$, $p > .05$; C_SM vs. C_SC: t (18) $= 1.72$, $p > .05$]. The overall difference between the Control condition and the three variable context groups taken together was statistically significant [t (38) $= 2.16$, $p < .05$].

These results are generally consistent with those of an earlier study

by Bevan and Avant involving the free recall of pictorial stimuli.[7] Both focal and background (contextual) stimuli were colored photographs of common objects mounted on a plain white field. When focal and background stimuli were randomly combined or when the background stimuli represented a broad category inclusive of the class of the focal stimuli (*e.g.* dogs vs. common mammals), the effect was a reliable reduction in the mean number of focal stimuli recalled as compared to a control group without context. When the focal and background stimuli were of the same class (*e.g.* different breeds of dogs), there was a slight, but statistically reliable, average increase in the number of stimuli recalled. When an associative-recall test was performed, the number of focal stimuli recalled in the presence of background items alone was directly related to thematic relevance.

On the other hand, the present results differ from the earlier findings of Pan, who reported that the presence of certain types of context words impeded learning but found that context words logically related to the stimulus constituted a slightly greater impediment to performance than logically unrelated words (13% vs. 8% fewer correct responses after the same number of presentations of the paired-associates test).[8] Pan did not account for the difference in level of interference between related and unrelated context words. How does one explain the present observation of a different relationship between thematic relevance and speed of learning? Postman and Phillips have suggested that when contextual words are high in meaningfulness, the likelihood is great that associations between contextual and response items will be formed which compete with response associations formed to the focal stimulus.[9] However, this explanation seems unlikely for the present data, since particular contextual items never appeared with more than one S-R pair, with the result that associations, once formed, should have converged rather than competed. Perhaps a more likely explanation is that contextual words, when unrelated to focal stimuli, served as distractions to *S* and hence retarded the formation of any association, while two closely related words formed a cohesive unit to which a mediating response could be more readily attached. Put differently, *S*s in the C_SR group, in their desire to identify the 'real' purpose of the ex-

[7] W. Bevan and L. L. Avant, The role of the thematic relationship between focal and contextual stimuli in recall, *J. verb. Learn. verb. Behav.*, 8, 1969, 185–190.

[8] Pan, *op. cit.*, 479–480.

[9] Postman and Phillips, *op. cit.*, 72–73.

periment, may have spent time trying to discover a logical relationship between context and stimulus items.

With respect to the general performance decrement for all three context groups, one interpretation—which is probably only partially adequate—involves the fact that *S*'s task was first to learn which of the two words was the stimulus item, before constructing an association between it and the response. While it is plausible that this would have resulted in an increase in trials to criterion, it seems likely that it would have added only about one trial to the initial learning scores, since a large number of *S*s spontaneously remarked during Trial 2 that they almost always could distinguish the proper stimulus word from the context word. It is not clear whether this was due to their recognizing that they had seen the stimulus word before or to their failing to recognize the contextual item, which was new to them.

Finally, inspection of the relearning scores indicates that by the time the initial criterion had been reached, the task had been thoroughly learned and immediate retest without the context word present resulted in only minimal loss: .9 trials in the case of the high-relevance group and .4 trials in the case of the low-relevance group. Pan similarly reported no difference between his high- and low-relationship groups in 24-hr. recall.

Performance with context and response words thematically related (the C_R groups). The lower half of Table II reveals the degree to which performance was influenced when the context word was thematically related to the response. When the thematic relationship was high, performance was greatly facilitated, by 62% [C_RH vs. C_RC: t (18) = 6.75, $p < .001$]; and when it was moderate, performance was better by 28% [C_RM vs. C_RC: t (18) = 2.58, $p < .02$]. These results are consistent with those of Pan. However, when the context and response words were randomly paired, the effect was to produce a decrement of 18% [C_RR vs. C_RC: t (18) = 1.12, $p > .05$].

In the case of facilitation, the reason seems readily apparent. The context word provides a clue of greater or lesser precision for the anticipation of the response word. The relearning scores clearly indicate that *S*s in the High and Medium context groups (C_RH and C_RM) did not acquire a strong association between the nonsense syllable and the response word during initial learning. When tested without context, they required virtually as many trials as *S*s in the Control groups to achieve errorless performance. However, when the

relation between context and response was minimal (C_RR), scores on relearning conformed closely to those of the C_S groups. Thus it seems obvious that *S* tended, when he could, to ignore the stimulus and seek out the context word as a cue to the response.

The matter of accounting for the decrement in initial performance of the random-relationship group is somewhat more difficult. Certainly, no portion of the decrement could have been caused by greater available time to distinguish between context and stimulus, since the latter was a CVC. One possibility is that the greater potency of high-frequency words as compared with low-association CVCs caused *Ss* to attempt first to form associations between the context and the response items. It is also conceivable—and this argument applies to the C_S groups as well—that interference may have been produced either by the simple physical presence of another word in the stimulus field in addition to the prescribed stimulus or by the requirement to pronounce the context word out loud. The Invariant-context groups were included to test this possiblity.

Performance with Invariant context (C_SI and C_RI). Inspection of Table II suggests a decrement associated with the addition of the irrelevant word, 'nutshell,' as an invariant contextual stimulus. The C_SI and C_RI groups both required 20% more trials to criterion, on the average, than their corresponding Controls. These decrements separately failed statistical confirmation [C_SI vs. C_SC: t (18) $= 1.50$, $p > .05$; C_RI vs. C_RC: t (18) $= 1.43$, $p > .05$]; however, when combined, t (38) $= 2.13$, $p < .05$.

It thus appears that the requirement to pronounce an additional word, either before or after the stimulus word, before giving the response may have been responsible for some, if not most, of the observed decrement in the other groups. If the relevant stimulus property for eliciting a mediating response is at least partly its auditory quality or pronounceableness, then the introduction of another word to be pronounced can reasonably be expected to have a distracting effect. (Visual distraction can be considered to be minimal here since, it will be recalled, the cue to pronounce 'nutshell' was a black line, not the word itself.) Since the percentage decrement for Invariant-context groups, at 20%, was very similar to that found for the other groups (except C_SR), it is tempting to invoke this vocal-auditory distraction hypothesis as the sole factor explaining decrement in the other groups. However, one would expect at least a portion of the impairment in the latter cases to be due to the

necessity that *S* learn a discrimination between stimulus and context. The additional 20% impairment found in C_SR could have been due to the other type of distraction discussed earlier, namely, *S*'s attempts to discern a logical connection between the two words which appeared together. This, furthermore, may well explain why Pan, who did not require word pronunciation, obtained a decrement in a situation comparable to our C_SR.

A logical correlate of the vocal-auditory distraction hypothesis is that a context word pronounced between the stimulus and response should be more detrimental than one pronounced before the stimulus. This hypothesis was not supported by an analysis of the errors on individual presentations for both Invariant groups, in either early, middle, or final stages of acquisition.

Summary

Ten groups of *S*s participated in a study of the role of verbal context in paired-associates learning. For five groups, English nouns served as stimuli, and the responses were low-association-value CVCs; in the other five groups, this relationship was reversed. When context words were varied and only randomly related to the focal stimulus, performance was significantly poorer than that of a control group presented no contextual stimuli. When context items were thematically related to stimuli, the context-induced decrement in performance was reduced. When they were related to response terms, identification of the response word was rapidly achieved. However, in this latter situation little if any learning occurred, since retraining to criterion without benefit of context revealed performance not reliably different from that of the control group. When the context was a single invariant word to be pronounced, performance was similar to that associated with other nonfacilitating contexts, strongly suggesting the operation of vocal or auditory distraction. Finally, learning without context appears not to have been reliably more difficult when stimuli were nouns and responses were CVCs than for the opposite condition; however, after criterion was reached, reversing the S-R relationship resulted in clearly better performance for *S*s then given CVCs as stimuli.

FORM PERCEPTION WITHOUT A RETINAL IMAGE

By Irvin Rock and Fred Halper,
Institute for Cognitive Studies, Rutgers University

One of the salient attributes of visual experience is extensity. Visual objects are extended in space; that is to say, the points making them up are located in separate places. It has always been assumed that the basis of visual extensity, and hence of visual form perception, is the retinal image, which captures the information about the spatial extensity of the object in such a way that the separate points of the object yield correspondingly separate points in the image. All theories of form perception start with the extended retinal image as the necessary first step.

However, the phenomenal location of a point with respect to the observer is not only a function of the location of its retinal image. The position of the eye must also be taken into account. With changes in eye position, a point in space which remains in a given phenomenal location stimulates the retina in varying locations; conversely, an image in a constant retinal location is perceived in varying phenomenal locations as the eyes move. In other words, there are two possible interpretations of the basis of extended visual form: (1) in terms of the extended retinal image per se, in which each point is a local sign of a different location, the perceived form being the result of a synthesis of all these locations; and (2) in terms of the phenomenally experienced direction of each point in the extended retinal image as that direction results from taking eye position (and possibly other information) into account, the perceived form being the result of a synthesis of all these directions. Ordinarily both interpretations are possible because the location of points in an extended image relative to one another will be paralleled by a corresponding phenomenal location of these points relative to one another. This is so because for any given position of the eye, the phenomenal direction of the entire array of retinal points is apparently simultaneously interpreted on the basis of that eye position.

* Received for publication June 1, 1969. This research was supported by a Public Health Service Research Scientist Award (5-K05-MH31361) to the first author. We wish to thank Alan Gilchrist for the construction of the apparatus employed.

Thus, for example, dots a and b ordinarily give rise to two corre-

$$\underset{a}{\cdot} \qquad \underset{b}{\cdot}$$

sponding images spatially separated on the retina and, if afterimages of a and b are formed, whatever the eye position, these images give rise to two corresponding phenomenal locations spatially separated, with a seen to the left of b. For different positions of the eyes, the radial or egocentric location of the pair a b changes, but their perceived separation and location with respect to one another does not change.

It is possible, however, to isolate retinal location from phenomenal location as the basis of form perception by presenting successively the stimuli that constitute the form. The experiments reported here attempted to achieve this isolation. The main idea was to have *O* view a moving spot, the path of which described a figure of a given shape. If *O* fixates and tracks the spot, and perceives the shape of the path, that perception can not be based on a retinal image of the path since no such retinal image is present; if, however, *O* fixates a stationary spot and perceives the shape of the path described by the moving spot (which does, thus, trace out a retinal image), the question at issue is whether that perception is facilitated by the presence of the retinal image; if, finally, *O* fixates the moving spot and perceives the shape of a 'path' described by a stationary spot, that perception must be based exclusively on the path of the retinal image—*i.e.* it cannot be based on changing phenomenal location since the stationary spot would be seen to remain in a constant phenomenal location.

Accordingly, a preliminary experiment required *O*s to note and then draw the shape of a path taken by a luminous spot in a darkened room under each of three conditions. In *Condition 1, O* tracked a single luminous spot with his eyes as it was moved by *E* along the perimeter of an irregular plywood shape that was not visible to *O*. In *Condition 2, O* fixated a stationary luminous spot centrally placed on the figure and noted the shape of the path taken by a second spot (of a different color) as it was moved around the perimeter. *Condition 3* was similar to Condition 2, but here *O* was required to track the moving spot and note the shape of the path made by the retinal displacement of the stationary spot. In Condition 1 the retinal locus of stimulation thus remained constant while changes in phenomenal location took place. If an extended image is a necessary factor in form perception, *O*s should have had no impression of a shape. The apprehension of shape in Condition 2 was a function of changes in

both retinal and phenomenal location. If an extended image is a factor in form perception, *Os* should have done substantially better in this condition. (Of course, it must *not* be the case that the moving spot is displacing so rapidly that its entire path becomes visible as a result of afterimagery, like the path of a cigarette moved very rapidly in a dark room. The shape of that path would be given by a 'simultaneous' image.) In Condition 3 only alterations in retinal locus occurred; there was no change in the phenomenal location of the stationary point being judged. If an image is a sufficient factor in form perception, the shape of the 'path' should have been perceived.

The *Os'* drawings under Conditions 1 and 2 were fairly accurate, and there was no evident difference in accuracy between the two conditions. On the other hand, most *Os* in Condition 3 were unable to produce a recognizable representation of the path. These results clearly argued in favor of phenomenal location as the critical factor. But one can criticize this experiment as not relevant to form perception: *Os* did not have the impression of seeing a *figure* of a given shape but rather of seeing a spot tracing out a *path*, which path had a given shape. Furthermore, the path can be held to be *inferred* rather than seen. We thought that this shortcoming was a function of the relatively long time required for the spot to traverse the entire figure. In the experiments described below, we sought to overcome this difficulty by employing a different method of successively exposing the parts of a figure.

Experiment I

Method. The three conditions of the preliminary experiment were tested using a different technique of exposing the figures. A narrow vertical slit, ⅛ in. wide × 3 in. high, was cut in a horizontally moving endless belt running over two cylinders. The belt moved over a luminous single-line figure approximately 2 in. long (see illustrations in Figs. 1 and 2) at about 3 in. per second. We were thus able to expose the entire figure at a high speed and fairly consistent rate while meeting the requirement that only a discrete luminous element be presented. The slit passed only once over the figure. *O* was about 2 ft. away.

In *Condition 1*, *O* was instructed to track the luminous element, and nothing else was visible. In *Condition 2*, *O* was instructed to fixate a stationary violet luminous spot ⅛ in. in diameter. The spot was placed on a transparent plastic strip centered with respect to the two ends of the figure revealed by the moving slit behind. In *Condition 3*, *O* was instructed to track the luminous element but to note the shape of the path of the violet spot. Each *O* was shown three figures, one for each of the three conditions. The order of presentation was randomly varied. A fine-discrimination recognition task employed the correct figure and nine similar alternatives arranged in a circular pattern. *O* was

required to identify the correct figure on the recognition chart after each figure was exposed through the slit.

Observers. Twelve undergraduates served as *O*s. They were naïve with regard to the purpose of the experiment.

Results. In Condition 1, eight *O*s were able to identify the shape correctly, and one *O* identified it correctly when given a second choice. In Condition 2, eight *O*s were able to identify the shape correctly. In Condition 3, only two *O*s reported seeing a path 'traversed' by the stationary luminous spot, and only one of these correctly identified the shape. It should be pointed out that in Condition 3 a correct identification required the selection from the recognition chart of an alternative that was the inverted mirror image of the path traversed by the exposed element. This was the case because the path of the image of the stationary spot over the retina is reversed with respect to the path of the image of the 'moving' element.

Discussion. It would seem that the apprehension of shape is to a great degree dependent upon information about the successive changes of phenomenal location of a single point. The wide difference in performance between Conditions 1 and 3 indicates the importance of the phenomenal factor. As a result of the factors responsible for position constancy, the retinally displacing spot in Condition 3 was not perceived to have moved and therefore yielded no impression of a path. It is noteworthy that in Condition 3 there was a tracing of an image across the retina although no shape was apprehended. The results of Condition 2 were consistent with this interpretation in that the addition of retinal changes of location to the phenomenal changes of location did not enhance performance.

To what extent did our improved method of presentation obviate the criticism that we were not investigating form perception? The answer is that it eliminated one criticism but not the other. With the new procedure, the *O*s in Conditions 1 and 2 did *perceive* the path of the 'moving' element of light; they did not have to infer it. But they still did not have the impression that they were looking at a *figure* the parts of which were being successively exposed. Therefore one might say we were here studying path perception rather than form perception, if the latter is defined in terms of the phenomenal experience one normally has in viewing a figure.

At about this time, a brief paper by Parks came to our attention.[1] In it Parks reported that when an outline figure is moved at an

[1] T. E. Parks, Post-retinal visual storage, this JOURNAL, 78, 1965, 145–147.

optimal speed behind a narrow, stationary slit, the entire figure is perceived, although somewhat foreshortened, an effect that had been studied by Zöllner and Helmholtz in the last century.[2] Parks concluded that the successive columns of stimulation projected onto the retina with no horizontal extent (beyond the width of the slit) are postretinally stored and then reconstructed spatiotemporally to yield a percept of the whole figure. There are certain similarities between the method employed by Parks and the method used here in Experiment I. In both cases, the parts of a figure successively came into view through a narrow slit. There are, however, certain differences, chief among which are (*a*) presentation of a luminous figure in a dark room vs. presentation of a line figure in the light and (*b*) movement of the slit over the figure vs. movement of the figure behind the slit.

As to the second difference, it seemed obvious that if the *moving-figure method* led to an impression of a figure, our *moving-slit method* should do so all the more easily and reliably.[3] The basis for this assertion is that with our method, the phenomenal locations relative to *O* of the parts of a figure are successively revealed whereas this is not the case with the moving-figure method. In fact, the mechanism underlying figure perception when a figure moves behind a stationary slit is not at all clear unless, as has been claimed in two recent papers, it is an artifact of eye movements resulting in a "painting" of an

[2] F. Zöllner, Ueber eine neue art anorthoskopischer Zerrbilder, *Ann. Phys. Pogg. Ann.*, 117, 1862, 477–484; H. von Helmholtz, *Handbuch der Physiologischen Optik*, transl. J. C. Southall, 3, 1962, 251. Recently Hochberg has reported on research in which *Os* were presented with a figure that moved behind a circular aperture. The aperture was large enough to permit simultaneous vision of the corners of the cross figures he employed, so that some partial image of the entire form was always present. The purpose of this research was to simulate the succession of foveally yielded impressions that must occur in fixating the various parts of an extended form in everyday life and to show that such a succession of impressions can be organized into a schema of the entire figure. This technique is analogous to the moving-figure method used by Zöllner, Helmholtz, Parks, and others. See J. E. Hochberg, In the mind's eye, in R. N. Haber (ed.), *Contemporary Theory and Research in Visual Perception*, 1968, 309–336.

[3] Following the publication by Parks, *op. cit.*, an investigation was reported in which the moving-slit method was used (although the reason for doing so was not the same as ours). See R. N. Haber and L. S. Nathanson, Post-retinal storage? Some further observations on Park's Camel as seen through the eye of a needle, *Percept. & Psychophys.*, 3, 1968, 349–355. Gibson also used this method in a test he devised for pilots: successive presentations of the parts of a figure are seen through a narrow slot that moves over the figure. In another test, a single moving spot describes a complex path. The *O*'s ability to 'visualize' the form or the path is determined by a recognition test containing alternative patterns. See J. J. Gibson (ed.), *Motion Picture Testing and Research*, 1947.

image of the figure on the retina as the figure is revealed through the slit.[4] In point of fact, when in preliminary work we tried out our method with our figure and *with lights on*, it did generally lead to an accurate impression of the figure and with greater reliability than did the moving-figure method. It was thus evident that the first difference, namely dark vs. normally illuminated room, was crucial to whether or not *O* has the phenomenal impression of a figure. We therefore decided to use our method of moving a slit over a figure under conditions of full illumination. The hope was that this would come closer to duplicating the experience one normally has in perceiving a figure despite the conditions of successive presentation.

Experiment II

Method. The method was like that of Experiment I except that the figures were drawn in ink on a white background rather than being luminous and that the room lights were on. Condition 3 was discontinued because it was clear from Experiment I that it did not lead to perception of a path and because the new procedure introduced to yield figure perception could not be expected to do so for an isolated stationary spot not seen within a slit.

Observers. Thirteen undergraduate students were used. They were naïve with regard to the purpose of the experiment.

Results. Twelve of the *O*s in Condition 1 saw a line figure, and six of these correctly identified it on the recognition chart; two more did so when given a second choice. In Condition 2, ten of the *O*s saw a line figure and five of these identified it correctly; two more did so on their second choice. The *O*s who did not see a line figure apparently did not see a spot traversing a path either; instead, they saw a spot moving up and down within the slit as the slit moved sideways.

Discussion. Virtually all *O*s perceived a line figure, and a good number did so with sufficient accuracy to identify the figure in the recognition test. The recognition test apparently was a difficult one, because in a control experiment where the entire figure was exposed for 4 sec., only nine (of all 13) *O*s identified it correctly on their first choice.[5] Again, there was no advantage provided by Condition 2,

[4] See S. M. Anstis and J. Atkinson, Distortions in moving figures viewed through a stationary slit, this Journal, 80, 1967, 572–585; Haber and Nathanson, *loc. cit.* However, we have reason to believe that this effect is not the result of retinal "painting."

[5] In the case of the moving-figure method, what is seen is generally a distorted version of the figure presented (compression or expansion). We have observed a very slight distortion (compression) with the moving-slit method but did not allow for it in the alternatives offered in the recognition test. This would of course make it somewhat more difficult to find the correct figure.

where an image of the figure was traced over the retina. The result to be emphasized is that of Condition 1, in which the prevailing experience was of a line figure, in contrast to the experience of only a moving point in Condition 1 of Experiment I.

This result implies form perception without a retinal image (other than a small foveal region), if it was indeed the case that *O* had succeeded in *tracking* the element of the figure revealed through the slit. If, however, despite instructions, *O* had held his eyes relatively still, an image of the figure would be imprinted over the retina. Two separate studies employing the moving-figure method have concluded that a figure is seen only if the eyes move at the same time, *i.e.* that the figure percept depends upon an extended image; in one of these studies the behavior of the eyes was noted in the moving-slit method as well, and it was concluded that only if the eyes remained more or less stationary and an image thereby traced over the retina, was there a genuine perception of a figure.[6] This would seem to be in direct contradiction of what we believed our *O*s were doing in Condition 1 of Experiments I and II. Therefore it was important to establish that our *O*s were indeed tracking the exposed element.

Experiment III

Method. The basic plan here was to require *O* to track the small element filling the width of a slit moving over a stationary figure as in Experiment II, while we obtained electroculogram recordings of the movements of his eyes.

Apparatus. Eye position was recorded with a Beckman Dynograph equipped with two Direct-Current Nystagmus Couplers (Type 9859), an event marker, and a 1-sec. interval timer. The *E*s also saw that Beckman miniature skin electrodes (10 mm. diameter) were applied with adhesive collars to (*a*) the temporal edge of the socket of each eye, (*b*) the vertical midline immediately above and below the right eye, and (*c*) as ground, to the lobe of each ear. Ocular activity along the horizontal and vertical axes was thereby recorded on independent channels.

A new apparatus for displaying the line figure was constructed. A slit 4 ¼ in. high × ⅛ in. wide was cut in the 1/16-in.-thick opaque plastic rim of a vertical cylindrical drum that was 24 in. in diameter and 6 ¼ in. high. The drum rotated on a central vertical shaft with speed adjustable by means of a Variac control. It either rotated in one direction or oscillated back and forth. Behind the rim of the drum was a stationary white stand, comparable in curvature to the plastic rim. On the front of each stand a ⅛-in.-thick single-line curved figure was drawn, 7¼ in. in length, and of the kind used in the preceding experiments. The height of each of the three stimulus figures employed did not exceed the height of the slit. The figures were constructed to have relatively

[6] Haber and Nathanson, *loc. cit.*

gradual slopes. In this way, only a single discrete element was revealed as the slit moved over the figure. An event marker recorded the onset and termination of the excursion of the slit across the figure by means of brush contacts on the inner plastic surface. The duration of each exposure was recorded by means of a constant-interval signal. For the single-exposure trials the figures were traversed at a rate of approximately 3 ½ in. per sec., yielding exposures of about 2 sec. per figure; for the oscillation trials the values were 7 in. per sec. and 1 sec. per figure respectively.

Procedure. *O* was seated in a chair 3 ft. from the apparatus; his head fixed by means of a bite plate. Following the application of electrodes, a 20-min. period of electrical stabilization elapsed before beginning the trials. *Single-exposure trials.* Prior to each exposure, the slit was in a stationary position revealing the beginning of the figure. *O*'s attention was directed to the slit and the single element visible therein. As each figure was first presented for one exposure only, this method provided *O* with an initial fixation point in order to facilitate accurate tracking once the slit began to move. *O* was instructed to follow, as closely as possible, what appeared in the slit as it moved from right to left. Immediately after the figure was traversed, *O* was asked to report what he 'saw.' Great care was taken to ascertain the precise meaning of *O*'s report. These reports could indicate (*a*) the perception of a spot which moved vertically up and down as the slit moved, or (*b*) the perception of a shape of a path made by a spot, or (*c*) the perception of a line figure. Only those *O*s who reported the perception of an extended line figure were presented with the recognition-test array from which they indicated their choice. This procedure was repeated for each of the three figures presented to *O*.

—*Oscillation trials.* Single-exposure trials often leave something to be desired as far as figure perception is concerned. For many *O*s there is an urge to continue looking at the figure being uncovered in order to be sure of the nature of their experience. There is little question that for many the impression of a figure improves with repeated exposure, as if the percept requires time to be constructed. For this reason, following the eye-position recording of the single exposure, two *O*s were permitted to view the figure as the slit oscillated back and forth over it, and as eye position was recorded. They were again told to track the element seen in the slit. Two additional *O*s were included who were not given the single exposure previously.

Observers. Seven undergraduates with normal vision were employed. They were naïve with regard to the purpose of the experiment and were paid for their participation.

Results: Single exposure. Considering first the eye-position data, all *O*s successfully tracked the element in its horizontal component of displacement. However some *O*s were better able than others to track in the vertical direction; Fig. 1 is an example of such fairly accurate tracking. The difference between horizontal and vertical tracking is understandable, because the slit was visible and moved at a fairly constant rate, so that *O* could anticipate where his eyes must be pointed in order to keep pace with the slit. The vertical

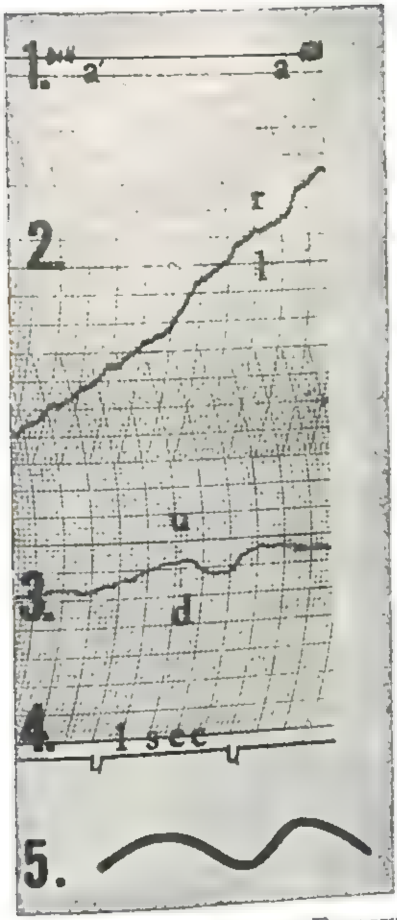


FIG. 1. EYE-POSITION DATA: SINGLE-EXPOSURE TRIALS
(1. Figure demarcation, beginning [a] and end [a′] of figure; 2. Horizontal axis, right [r] and left [l]; slope of recording attests to continuous tracking movement of eyes as slit moves right to left; 3. Vertical axis, up [u] and down [d]; since 2 indicates eyes keep pace with slit, this recording should mirror shape of figure if eyes track accurately in vertical direction; recording shows quite accurate tracking; 4. Interval marker, 1 sec.; 5. Figure presented)

displacement of the element, on the other hand, could not be anticipated, and when the figures sloped upward or downward, the change was quite rapid.

The crucial question was whether the *O*s perceived a line figure while tracking. Eleven of the figures presented (of a total of 15; five *O*s, three figures each) were perceived as line figures. (In the few cases where a line figure was not perceived, *O* generally perceived an element moving up and down the slit.) In six of these cases, both the vertical and horizontal components of the eye recording reflected

accurate tracking, thus demonstrating form perception in the absence of a retinal image. In the remaining five instances of figure perception, the records revealed accurate tracking in the horizontal direction but not with regard to vertical changes of the element. Finally, when both horizontal and vertical tracking were accurate, correct recognition occurred in four out of six instances. In the cases of accurate tracking in only the horizontal direction, two of the five figures were correctly identified. The failures of recognition are probably attributable to the high degree of similarity among alternatives.

—*Oscillation exposure.* All *O*s accurately tracked the horizontal component of the element's displacement, but few did so for the vertical component. With the speed of the slit even greater in this condition and the added strain of continual tracking, the failure of 'vertical' tracking is understandable. Fig. 2 is a representative record taken during oscillation. Figure perception was excellent under the oscillation procedure. Of 12 figures presented (four *O*s, three figures each), 11 were seen as line figures and 10 of these were correctly identified.

Discussion. In the single-exposure procedure, those instances in which the *O*s tracked accurately in both the vertical and horizontal direction *and* perceived figures which they could later identify, clearly demonstrate visual form perception with no more of an image than a small foveal region. But we feel that each instance of figure perception (in which all *O*s tracked accurately in the horizontal direction) is also clear evidence of form perception *not* based on a retinal image. In these cases the only variation in retinal stimulation was confined to the columnlike projection of the slit. One can hardly account for the perception of the figures we employed on the basis of an image of a small spot moving up and down a narrow column of retina. The success of horizontal tracking, together with the excellence of figure perception under the oscillation procedure, is further evidence that form perception is correlated not with the shape of the retinal image but with the phenomenal locations of the displacing element.

Simulation Method

It is possible to simulate a line figure being uncovered, since the stimulus conditions seem to consist of a small element that completely fills the width of the slit and moves up and down as the slit moves

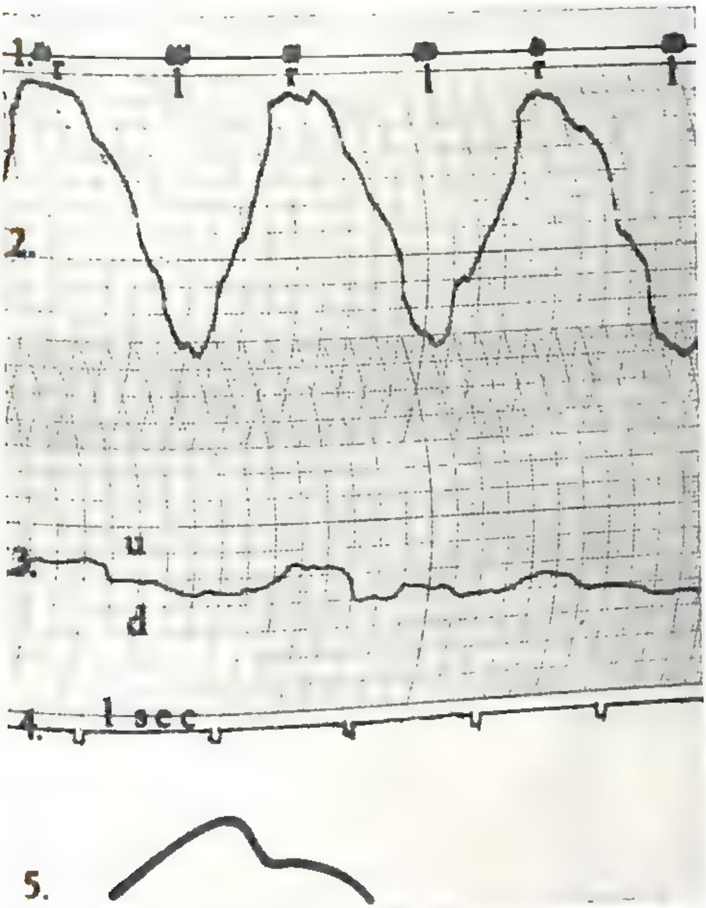


FIG. 2. EYE-POSITION DATA: OSCILLATION TRIALS
(1. Figure demarcation, right [r] and left [l] of figure; 2. Horizontal axis, right [r] and left [l]; recording shows several back-and-forth movements of eyes keeping pace with slit as it oscillates; 3. Vertical axis, up [u] and down [d]; recording indicates little if any tracking in vertical direction; 4. Interval marker, 1 sec.; 5. Figure presented)

back and forth. We achieved this by means of a thin wire, perpendicular to the surface containing the slit, on the side away from *O*. A small piece of black cardboard was affixed to the end of the wire directly behind the slit so as to approximate the way a segment of an actual figure would look when seen through the slit. As the slit moved back and forth, the wire went with it but an unseen part of the wire rode over a cam, or cutout, having the shape of the figure being simulated. Thus the element seen through the slit moved up and down as the slit moved sideways. Although it would be possible to arrange for the slope of the element to change—as it would if a

figure were being revealed—we found that it was not necessary to do so.

When the surface containing the slit was moved at the optimum rate, the impression of a line figure was as vivid as when an actual figure is being uncovered. Interestingly enough, the phenomenal slope of the element was congruent with the part of the figure being perceived at any moment although, in fact, the slope of the element remained constant under the conditions employed. Needless to say, the simulation method also failed to yield figure perception in a dark-field procedure.

Discussion

To what extent does the kind of figure perception studied in Experiments II and III typify ordinary form perception? Some might be inclined to say that we were studying form apprehension via inference rather than perception of form because only a small segment of the figure was seen at any one moment. Actually this would be an incorrect statement, confusing the physical state of affairs with the phenomenal. Physically, it is true that only a small element of the figure was present at one time, but there are many examples in perception of temporal integration, *e.g.* hearing a melody or experiencing a solid three-dimensional object based on the rotation of its two-dimensional shadow. No one would consider these experiences nonsensory. At a sufficiently rapid rate of exposure, our *O*s experienced a visual form and the experience was essentially indistinguishable from ordinary form perception, except that *O*s may have been aware the entire figure was not visible at any single moment.

As far as explanation in terms of a "painting" of an image over the retina is concerned, our eye-movement recordings would seem to be unassailable evidence against it. This finding is in direct contradiction of the report by Haber and Nathanson that "the complete percept disappeared" when their *O*s attempted to fixate a slit moving over a figure but reappeared when their *O*s' eyes remained fixated on some part of the figure.[7] They further stated that their *O*s found it difficult to track the moving slit and, unless instructed to do so, tended to hold their eyes still. They imply that an *O* who is tracking can correctly *judge* the shape of the total figure by a process of integrating "successive incomplete percepts" if the movement is slow, but that such judgment is clearly different from the genuine form

[7] Haber and Nathanson, *loc. cit.*

perception that occurs if a "painting" of the image is possible when the eyes are stationary. Our results on tracking indicate that an extended image is not necessary for perceiving form; the negative outcome of Condition 3, Experiment I, further indicates that an extended image is not a sufficient basis for form perception. Still, the possibility exists that an extended image may contribute something to the accuracy or vividness of the percept when the conditions necessary for figure perception are present. *O*s have occasionally commented that they see a figure more readily and more vividly if they hold their eyes still than if they track the moving slit. This was particularly evident using an apparatus constructed after the completion of the experiments described in this paper. In this apparatus, black line figures are seen on a brightly illuminated milk-glass panel. It is quite possible that these *O*s' improved percepts were based on an afterimage of the sharply contrasting element that was traced over the retina when the eyes were stationary. The speed of the moving slit is obviously a factor here. Such an afterimage would of course provide a more or less simultaneous image of the entire figure and, therefore, the improved percept would be expected. Haber and Nathanson's finding may be based on such afterimagery.

It is also possible that the eyes may behave in a particular fashion as a *consequence* of perceiving a form. Consider the case of a moving figure behind a stationary slit. If, for whatever reason, *O* does see a form moving through the opening, it is natural for him to move his eyes in order to remain fixated on a region of the figure. This is what he would do if the entire figure were visible during its movement. In the case of a slit moving over a figure, it is only by holding his eyes still that *O* can continue fixating a region of the figure. Thus the behavior of the eyes in these situations may be an effect of rather than a cause of figure perception.

In this connection it is interesting to consider the failure to obtain genuine figure perception in Experiment I, where luminous figures were seen in a darkened room. If retinal "painting" explains figure perception through a slit, why did it fail in this case? If anything, it would be easier to hold the eyes still during this procedure, since the entire moving surround was not visible. Furthermore, in two of our experiments, instructions to hold the eyes still (Condition 2) never resulted in improved perception of the figure despite the "painting" of an image thereby introduced, and in those experiments which included Condition 3—a pure test of the role of a tracing of an image over the retina—a line figure was never perceived. Condition 3

duplicates what happens all the time in daily life when we move our eyes. The images of all things sweep across the retina. Why does this never lead to an impression of the shapes of the paths taken by these moving images? Then too, when O views a figure moving behind a narrow slit, the vertical component of the "painting" results from the up-down displacement of the figure contour and the left-right component from the alleged horizontal eye movements. Suppose both components are generated by eye movements, by instructing O to move his eye along a curved path. In that case it is no longer necessary to move the figure, and a segment of it can remain in view while O moves his eyes. Clearly this procedure will not produce a figure percept, and one wonders how the retinal-"painting" advocate would account for the failure.

How do we account for the failure to obtain figure perception under the dark-field procedure of Experiment I? We believe something more is involved in figure perception through a slit than the presentation of a succession of phenomenal directions. We have reason to believe that what is crucial is the impression that one is viewing through a slit an object which is being successively exposed; and we will present evidence in support of this claim in a subsequent paper. This impression is not possible under a dark-field procedure because the slit is not seen per se, only successive portions of the luminous figure. That the difference between dark and light procedures is not merely the presence in the light of cues such as the microstructure in the figure or the paper being uncovered is evident from the fact that we have been able to obtain figure perception with a simulation method. Assuming we are correct in ruling out retinal "painting" as an explanation, the simulation method makes it dramatically clear that figure perception through a narrow slit is the result of a mental construction. A line figure behind the slit is a parsimonious 'solution' to the problem of what these unusual stimulus conditions represent in the world. The necessity of seeing the element fill the slit, or, otherwise expressed, of seeing the figure being uncovered by a slit, suggests a cognitive factor in the perceptual process. It should, however, be noted that not all Os see a line figure even under these conditions, nor does a given O always see a line figure. The stimulus is essentially ambiguous, so that despite a preference for one 'solution,' another (seeing an element moving up and down the slit) is also possible and does occur.

There are other perceptual phenomena the stimulus conditions of which have been thought to depend on the state of affairs on the

retina. For example, stroboscopic movement had been thought to result from successive stimulation of separate retinal loci. Grouping by proximity had been thought to depend upon the relative proximity of the retinal images of the stimuli entering into the grouping. Change in appearance of disoriented figures had been thought to be the result of altering the orientation of the retinal image of the figure. In all these cases, there is now evidence that it is not the state of affairs on the retina that is relevant, but rather the phenomenal state of affairs: the separately perceived location of the two stimuli in stroboscopic movement,[8] the perceived proximity of stimuli in grouping,[9] and the perceived change in orientation of disoriented figures.[10] Now we see that visual form per se is a function not of the retinal location but of the phenomenal location of the elements of the figure.

If retinal extensity, either of the typical simultaneous kind provided by an image or of the successive kind that occurs in retinal "painting," is not necessary for phenomenal extensity, the question arises as to the role of the retinal image. A suggested answer is that an extended image makes possible the *simultaneous* phenomenal location of all parts of a figure. This has two ramifications. First, we can perceive an entire figure—or, for that matter, an array of figures—in a single glance. Second, the impression of form is probably more accurate and certain than can be the case when successive presentation is used, as in the experiments reported here, because there is no need to rely on memory. A retinal image permits more accurate location of points with respect to one another because all points are simultaneously given. That it is the simultaneity rather than the extensity of the image that yields this probable benefit is borne out by the lack of any difference in our experiments between Conditions 1 and 2. The reliance on immediate memory in successive presentation undoubtedly explains why too slow a rate of movement of the slit fails to produce a genuine perceptual impression of a line figure.

[8] I. Rock and S. Ebenholtz, Stroboscopic movement based on change of phenomenal location rather than retinal location, this JOURNAL, 75, 1962, 193–207.

[9] I. Rock and L. Brosgole, Grouping based on phenomenal proximity, *J. exp. Psychol.*, 67, 1964, 531–538.

[10] I. Rock, The orientation of forms on the retina and in the environment, this JOURNAL, 69, 1956, 513–528; I. Rock and W. Heimer, The effect of retinal and phenomenal orientation on the perception of form, this JOURNAL, 70, 1957, 493–511; I. Rock and R. Leaman, An experimental analysis of visual symmetry, *Acta Psychol.*, 21, 1963, 171–183.

Summary

Experiments were performed to isolate the phenomenal from the retinal location of a set of points as the basis of form perception. In Experiment I, *Os* either tracked a 'moving' luminous spot or fixated a stationary spot while observing the path of the moving one. They perceived the path as accurately in the first case as the second, despite the absence of an extended retinal image, and presumably did so on the basis of information via eye position concerning the varying phenomenal locations of the moving spot. It was also shown that variation of retinal location in the absence of variation of phenomenal location does not yield an impression of a path. Thus retinal location is neither a necessary nor a sufficient basis of form perception.

A limitation of Experiment I was that a path was seen rather than a line figure. With the room lights on in Experiment II, however, *Os* viewing a line figure successively uncovered by a narrow moving slit did perceive a figure even when tracking the small region visible through the slit. The effect was also obtained by a simulation method in which a small element filled the slit and moved up and down as the slit moved back and forth. Recordings of eye movements established, in Experiment III, that *Os* were tracking when instructed to do so and, thus, that form perception without an extended retinal image is possible. Since the same technique failed to produce a line-figure percept when the figure being uncovered (or simulated) was luminous and viewed in the dark, it was concluded that a necessary condition is the impression of something being uncovered behind a slit. Finally, it was suggested that the advantage of an extended retinal image lies in its simultaneity rather than its extensity.

THE DIMENSIONALITY OF SIMILARITY JUDGMENTS

By LENNART SJÖBERG, University of Uppsala, Sweden

The present study investigated the dimensionality of judgments of similarity and the generality of certain relations among unidimensional scaling methods—relations that have been found to hold in studies of variables characterizing single objects rather than pairs of objects. Similarity is, of course, a variable characterizing pairs of objects; and the question of the dimensionality of similarity judgments involves at least two, quite different, aspects. The basic problem is whether it is at all possible to measure the similarity (of pairs of objects) for one individual on one occasion. Given that such measurement is possible, we may inquire into the structure of inter- and intraindividual variability of similarity judgments. Such structures may be described in spatial terms, and the results of such description are called second-order similarity spaces in this paper. These second-order spaces, then, involve a set of similarity scales generated by different individuals, or by one individual on different occasions, or both. Each such scale can be analyzed for a spatial structure of the stimulus objects. Such an analysis would give a first-order similarity space. Noting that the questions of second-order and first-order similarity spaces as conceived in this paper arise only if the basic measurement of similarity is possible, we turn now to a more detailed consideration of the measurement problem.

Similarity measurement. The possibility of measuring similarity has seldom been questioned, although little data directly bearing on the issue has been published. The most common way of attempting to measure similarity seems to be to use category judgments of the similarity of pairs of stimuli. For example, both Attneave and Shepard have discussed the properties of similarities necessary for a spatial representation of stimulus objects, both assuming that a

* Received for publication November 8, 1968. The study reported was supported by a grant from the Swedish Council for Social Science Research. The author expresses his indebtedness to Mr. Lars Bergman for experimental and computational assistance.

unidimensional similarity scale was available.[1] Tucker and Messick have suggested a method for analyzing similarity ratings[2] that is mentioned later in this paper; here it is sufficient to note that their method also assumed the existence of a unidimensional scale of similarity for each individual. Indeed, it is probably fair to say that the great majority of both theoretical and empirical work on similarity has utilized category judgments, tacitly assuming these to yield meaningful measures of similarity. However, similarity methods explicitly recognize the possibility of multidimensionality of objects. Therefore, it appears somewhat unsafe to assume unidimensionality of similarity judgments in this sense. The assumption should be subjected to empirical test. True, evidence may be presented for the thesis that unidimensional scales can often be successfully constructed,[3] but all the studies cited in favor of this assertion were concerned with scaling single objects, not pairs of objects as is the case in similarity studies.

In previous work, three criteria of unidimensionality have been used: tests of transitivity of paired comparisons, tests of fit of unidimensional scaling models, and comparisons of scale values resulting from the use of different response modes, such as magnitude estimation and category judgments.[4] The present study made use of all these criteria. The analysis of transitivity may easily be performed at the individual level and was, hence, the most directly relevant one. To study the fit of scaling models and to compare response modes requires more stable data, in the present study achieved by means of pooling data for a group of subjects. Consequently, evidence based on fit and comparison of response modes may here be less informative than data on transitivity.

Second-order similarity spaces. As mentioned above, second-order spaces result from variation between individuals, or occasions, or both; and they refer to pairs of objects. Tucker and Messick suggested the use of component analysis of cross products for studying the second-order spaces generated by judgments of similarity or difference. They interpreted the components as "points of view"

[1] F. Attneave, Dimensions of similarity, this JOURNAL, 63, 1950, 516–556; R. N. Shepard, Attention and the metric structure of the stimulus space, *J. math. Psychol.*, 1, 1964, 54–87.

[2] L. R. Tucker and S. Messick, An individual differences model for multidimensional scaling, *Psychometrika*, 28, 1963, 333–367.

[3] L. Sjöberg, The dimensionality paradox in comparative judgment: A resolution, *Scand. J. Psychol.*, 9, 1968, 97–108.

[4] Sjöberg, *loc. cit.*; L. Sjöberg, Unidimensional scaling of multidimensional facial expressions, *J. exp. Psychol.*, 78, 1968, 429–435.

regarding the similarities studied.[5] The present study attempted to investigate the second-order spaces of similarity judgments in the spirit of Tucker and Messick, although factor analysis of correlations was here used rather than component analysis of cross products and the factors were here interpreted in a conventional way rather than using the Tucker-Messick method of "idealized individuals."

Another way of studying the covariation space is by means of the variability of comparative judgments. The usual formula for the variance of a difference illustrates how such variabilities are contingent on correlations:

$$S_{i-j}^2 = S_i^2 + S_j^2 - 2S_iS_jr_{ij},$$

with conventional notation. We see that when r_{ij} increases, S_{i-j} decreases. If the two objects compared are interpreted geometrically as vectors, we may interpret S_{i-j} as the distance between the termini of the vectors. This interpretation suggests that *SD*s of differences may be analyzed by any of a number of multidimensional procedures that are available for analyzing distances.[6]

Scaling methods for paired-comparisons data have usually assumed correlations to be equal or zero.[7] In such cases, the dispersions of comparative judgments yield no information about dimensionality. However, the availability of methods not making such strong assumptions regarding correlations[8] now makes it possible to study the possibly multidimensional structure of these dispersion terms. An attempt is also made in the present study to investigate the possibility that dispersions resulting from magnitude estimations also carry multidimensional information. A weak trend of such an effect was found in an earlier study.[9] The Thurstonian method as applied in the present study does not allow separation of inter- and intraindividual variability. However, any interesting structure in the pooled variability would make it important to try to find the source to which it should be attributed,

[5] Tucker and Messick, *loc. cit.*

[6] W. S. Torgerson, Multidimensional scaling of similarity, *Psychometrika*, 30, 1965, 379–393.

[7] W. S. Togerson, *Theory and Methods of Scaling*, 1958; R. D. Bock and L. V. Jones, *The Measurement and Prediction of Judgment and Choice*, 1968.

[8] L. Sjöberg, The law of comparative judgment: A case not assuming equal variances and covariances, *Scand. J. Psychol.*, 3, 1962, 219–225; L. Sjöberg, A study of four methods for scaling paired comparisons data, *Scand. J. Psychol.*, 6, 1965, 173–185; L. Sjöberg, Successive intervals scaling of paired comparisons, *Psychometrika*, 32, 1967, 297–308.

[9] Sjöberg, *op. cit.*, *J. exp. Psychol.*, 78, 1968, 429–435.

and the design of the magnitude estimations was set up to make such an investigation possible. In sum, the main question about second-order spaces was whether meaningful second-order spaces would indeed appear in dispersions of comparative judgments. Two additional aspects of interest were (*a*) the structure in correlations among similarity judgments and (*b*) the separation of inter- and intraindividual variability.

Comparison of scaling methods. A major purpose of the present study was to investigate relations among scaling methods. It was considered of interest to attend to this problem in the present context because almost all earlier work on relations among scaling methods has been concerned with variables characterizing single objects. Work by Ekman and Künnapas and others has repeatedly demonstrated a curvilinear relation between Thurstonian paired-comparisons scales and scales resulting from the application of so-called direct methods (*e.g.* magnitude and ratio estimation).[10] The author's method of difference ratings, which is a variation on Thurstonian methods using less strict assumptions regarding certain dispersion terms, has, however, proved to be approximately linearly related to direct methods.[11] These results were followed up in the present study, which also investigated the relation between conventional category judgments of similarity and magnitude scales of similarity.

Classification of similarities. Fifteen pairs of pictures showing different emotions in different intensities were used as stimuli for the similarity-scaling by the different methods. It was hypothesized that three different kinds of similarity would emerge, either in intransitivity of paired comparisons of similarity, or in second-order similarity spaces, or both. The kinds of similarity considered were: similarity between pictures expressing the same emotion in different intensities, similarity between pictures expressing different emotions of approximately equal intensity, and similarity between pictures expressing different emotions in different intensities. The first kind of similarity hypothesized was termed quantitative, the second qualitative, and the third both qualitative and quantitative, or mixed.

[10] Extensive references may be found in: S. S. Stevens, A metric for the social consensus, *Science*, 151, 1966, 530–541; and Sjöberg, *op. cit.*, *Scand. J. Psychol.*, 9, 1968, 97–108.

[11] Sjöberg, *op. cit.*, *Psychometrika*, 32, 1967, 297–308; Sjöberg, *op. cit.*, *Scand. J. Psychol.*, 9, 1968, 97–108; Sjöberg, *op. cit.*, *J. exp. Psychol.*, 78, 1968, 429–435.

METHOD

Stimuli. Six pictures from the Lightfoot series[12] were selected: Nos. 14, 35, 33, 52, 47, and 22. These pictures had all been used in an earlier study, and scale values of emotional intensity were available.[13] The pictures had also given evidence of being relatively pure as to emotional content in a pilot study. They formed three different clusters of emotions, with two degrees of intensity in each cluster, as illustrated in Table I. The table also gives the letter designations used in this report for the pictures. It should be noted that the term 'stimulus' in this study does not denote single pictures, but pairs of pictures. Using all possible pairs of the six pictures, there were 15 stimuli (15 pairs). The stimuli were named according to the two pictures composing them. For example, Stimulus AC consists of the two pictures A and C. The paired-comparisons task involved judging pairs of stimuli, *i.e.* pairs of pairs of pictures (method of tetrads). There were 105 such different pairs.

Subjects. The *S*s were, with a few exceptions, psychology students, 112 in all: 79 women and 33 men.

Stimulus presentation. The experiment was conducted in a seminar room. A group of 5–10 *S*s participated in a given session. The pictures were projected on a screen in front of the *S*s by means of two Leitz projectors. Each picture contained a pair of facial expressions. Its projected size was .70 × .91 m. The distance between the two stimuli was .45 m. In each stimulus, the two faces were separated by .01 m. *S*s were seated 3–4.5 m. from the screen.

Procedure. Three kinds of rating methods were employed by each *S*: paired comparisons, category ratings, and magnitude estimation. The methods were always employed in the order mentioned. Two or three different random orders of presentation were employed, equally often forward and backward. Each *S* gave only one rating for each pair of stimuli in paired comparisons and for each stimulus in category judgment. In the case of magnitude estimation, four different standards were employed. The *S*s were divided randomly into four groups, each group using only one of the standards and giving three estimates of each variable stimulus. Each stimulus occurred in two different versions according to the position of the two faces included. These two versions were used equally often. In addition, positions within pairs of stimuli were also rotated over *S*s. Responses were recorded in booklets with five responses on each page. About 30 sec. were allowed for each response. The *S*s who did not give a response in time were allowed to complete their response

TABLE I

SELECTED PICTURES

Emotion	Intensity: Low	Intensity: High
Happiness	A = No. 14†	B = No. 35
Depression	C = No. 33	D = No. 52
Anger	E = No. 47	F = No. 22

† Lightfoot series number.

[12] T. Engen, N. Levy, and H. Schlosberg, A new series of facial expressions, *Amer. Psychol.*, 12, 1957, 264–266.

[13] Sjöberg, *op. cit.*, *J. exp. Psychol.*, 78, 1968, 429–435.

sheets before the group moved on to the next rating method. In such cases, *E* exposed the required pictures once more.

—*Paired comparisons.* Seven rating categories were used, plus the category of equality. The scale was presented as a string of numbers (765432101234567), on which a rating of 0 represented no difference in similarity ("the same similarity"). Ratings to the left of 0 indicated the pictures of the left pair to be more similar than those of the right pair, and ratings to the right indicated the opposite. Larger category numbers indicated larger differences in similarity. Category 7 was defined verbally as indicating "a very large difference in similarity."

The *Ss* were first instructed that their task was to compare similarities of pairs of facial expressions. They were asked to decide initially which pair showed the largest similarity and then to rate the size of difference of the two similarities. It was stressed that there were no right or wrong answers but that *E* was concerned with their personal opinions about the similarities in question. Four practice trials were then given, utilizing schematic drawings of faces. In three of these trials, the two expressions in one pair were identical, and in a fourth trial, the pairs were identical except for position. Then 10 additional practice trials were given, using pictures from the Lightfoot series (not the ones used in the main experiment). These trials gave *Ss* experience with the range of similarities used in the main experiment. Finally, it was strongly stressed that *Ss* were to differentiate between similarities within pairs and differences between similarities. The *Ss* were encouraged to ask about anything unclear. Apparently the distinction between the two kinds of relation was easily grasped. All instructions were given to *S* in writing and also read aloud by E.

—*Category judgments.* Seven categories were used, defined by numerical labels. The meaning of the scale was explained by means of a few examples. A rating of 1 was to be given to "a very small similarity," 7 to "a very large similarity." Each stimulus was presented only once.

—*Magnitude estimation.* The standard similarity was assigned the number 10. This number was also written in the lower edge of the standard stimulus. The variable stimulus was to be assigned a number reflecting the relation between the two similarities: an intensity half of the standard would be given the number 5, one twice as large the number 20, and so forth. Any numbers could be used as responses. The standard appeared at random in both positions and was shown anew for each variable stimulus. Stimuli AB, AC, BC, and BD were used as standards. Each *S* was exposed only to one standard and gave three estimations for each of the variable stimuli.

The entire experiment lasted about 130 min. During this time, *S* was required to make a total of 162 responses, not counting practice trials.

Results

Circular triads. The frequency of circular triads in paired-comparisons data was rather high. The mean number of such triads was 16.7, or 11.9% of the maximum possible number of circular triads. (An upper limit to the mean number of circular triads to be expected if equality judgments had not been permitted was esti-

mated at 43.7 triads. This estimate was based on the assumption that 30.8% of the triads containing at least one equality judgment would have turned out to be circular. The 30.8% is the probability that a maximally intransitive *S* would produce a circular triad, *i.e.* the maximum possible number of circular triads divided by the total number of triads.) The distribution of circular triads across *Ss* was very skew, a few *Ss* contributing heavily to the high arithmetic mean. The distribution across triads was also skew, but with no triad being circular for more than 19 *Ss*. Thus, if consistent intransitivities exist, they tended to be different for different *Ss*. The rather high frequency of intransitivities could, of course, also depend on low reliability of paired-comparisons judgments. This alternative could not be checked in the present study, since each pair of stimuli was exposed only once to each *S*.

To investigate whether the intransitivities were dependent on the existence of different kinds of similarity, the following two analyses were made. Stimuli were first divided into three groups: quantitative, qualitative, and mixed similarities. Each triad of stimuli was then classified as simple or complex, depending on whether all the three pairs involved came from the same group of similarities or not. The mean number of circular triads for simple triads was 3.90 and for complex triads 4.13, a very slight difference. Thus, using this classification of similarities, no systematic trend could be observed. A second analysis utilized a different classification of similarities. Each stimulus was classified according to the pair of emotions compared: happiness-anger, happiness-depression, or depression-anger. Within-emotion stimuli (AB, CD, and EF) were excluded from this analysis. Again, there was no trend for complex triads to be more intransitive than simple triads. The mean number of intransitivities was 4.33 for simple triads and 3.38 for complex triads. In passing, it may be noted that deleting the three within-emotions stimuli from the analysis produced a mean rate of circular triads of 9.6% as compared with 11.9% when they were present. The frequency of equality statements was 7.6%.

Scaling similarity. A summary of results of application of various scaling procedures is given in Table II.

—*Thurstonian scaling.* Paired-comparisons data were analyzed according to a procedure described by Sjöberg, and the category judgments were analyzed according to the method of Diederich,

TABLE II

SIMILARITY-SCALE VALUES

Stimulus	Paired comparisons	Category judgments†	Magnitude estimation‡ Standard AB	Standard AC	Standard BC	Standard BD
AB	1.24	1.00	1.00	1.00	1.00	1.00
AC	1.11	.80	.80	.67	.88	1.00
AD	.74	.62	.50	.53	.65	.75
AE	1.01	.77	.80	.67	.88	.85
AF	.55	.54	.40	.33	.59	.60
BC	.41	.58	.50	.33	.59	.60
BD	.19	.48	.30	.27	.47	.50
BE	.67	.70	.70	.53	.71	.65
BF	.00	.00	.20	.13	.35	.35
CD	1.40	.94	1.10	1.00	1.00	1.25
CE	1.26	.83	1.00	.80	.94	1.00
CF	1.18	.86	.90	.60	.88	1.00
DE	1.31	.90	1.00	.80	.94	1.00
DF	1.23	.92	1.00	.93	1.06	1.00
EF	1.19	.90	1.00	.93	.94	1.00

† Scaled according to the Thurstonian successive-intervals method.
‡ Medians, transformed to yield a value of 1.00 for Stimulus AB.

Messick, and Tucker.[14] The square root of the mean square deviation between empirical and reproduced proportions was .043 for paired comparisons and .032 for category judgments. Both goodness-of-fit values were somewhat low. The category boundaries were spaced in the usual manner: larger differences between more extreme categories. The relation between the two Thurstonian scales was nonlinear, as may be seen in Fig. 1. The result is the same as that found in a previous study.[15]

—*Magnitude estimation.* The median ratings under the four standards constitute four magnitude scales which may be compared in Fig. 2. Once again, data were somewhat variable, but the four scales may be said to be approximately linearly related to each other. There were some clear instances of nonzero intercepts in plots where the two standards differed greatly as to overall level. This finding replicates the results of the earlier facial-expressions study,[16] in which positive intercepts were also found if the larger standard was plotted on the y-axis. For further analysis, the four magnitude scales were then averaged.

—*Relations among scales.* Relations between Thurstonian and magnitude scales may be studied in Fig. 3. While the paired-com-

[14] Sjöberg, *op. cit.*, *Psychometrika*, 32, 1967, 297–308; G. W. Diederich, S. J. Messick, and L. R. Tucker, A general least squares solution for successive intervals, *Psychometrika*, 22, 1957, 159–173.
[15] Sjöberg, *op. cit.*, *J. exp. Psychol.*, 78, 1968, 429–435.
[16] Sjöberg, *op. cit.*, *J. exp. Psychol.*, 78, 1968, 429–435.

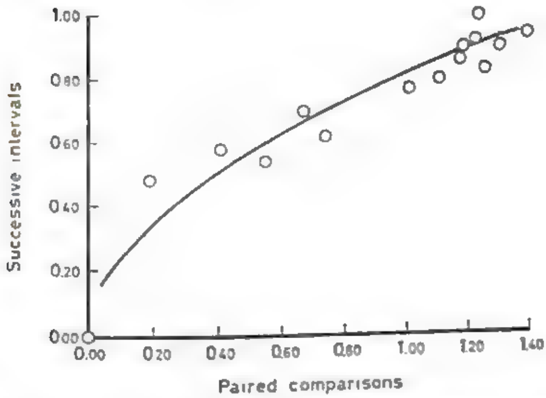


FIG. 1. RELATION BETWEEN PAIRED-COMPARISONS AND SUCCESSIVE-INTERVALS SCALES
(Regression line fitted by eye)

parisons scale and the magnitude scale were approximately linearly related, the Thurstonian category scale was a nonlinear function of the magnitude scale. This is in accordance with earlier findings. Finally, it may be of some interest to compare arithmetic means of category judgments to the magnitude scale. The plot in Fig. 4 shows good agreement between the two scales. It may be noted that both fit of scaling methods and relations between scales indicated more variability in data than is usual with variables characterizing single objects.

Factor analysis of category judgments. A factor analysis was carried out on category judgments using squared multiple correlations as communality estimates and factoring according to the method of principal factors. Inspection of roots showed that two or three factors should be considered. The solution was rotated to orthogonal simple structure (varimax). Table III gives the factor loadings for the three factors.

Interpretation of the three factors is rather straightforward if within-emotion pairs are disregarded. The three kinds of similarity hypothesized did not emerge. Instead, the factors pertained to depression-anger, happiness-anger, and happiness-depression comparisons. The factors seem to represent differing points of view as to the similarity between the three kinds of emotional content as sampled in pairs. Although the within-emotion pairs make this interpretation somewhat doubtful, it may still be useful as a summary of the results regarding between-emotion pairs.

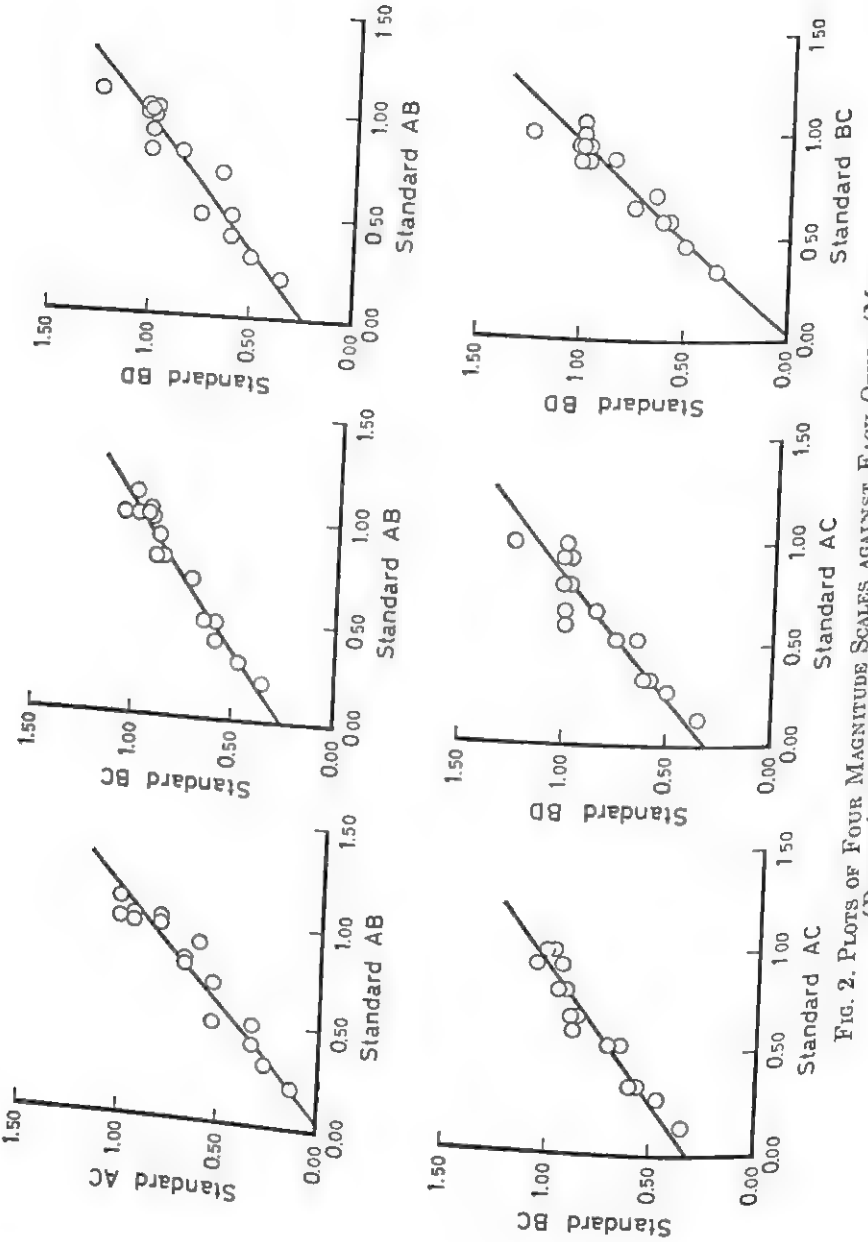


FIG. 2. PLOTS OF FOUR MAGNITUDE SCALES AGAINST EACH OTHER (MEDIANS)
(Regression lines fitted with method of least squares)

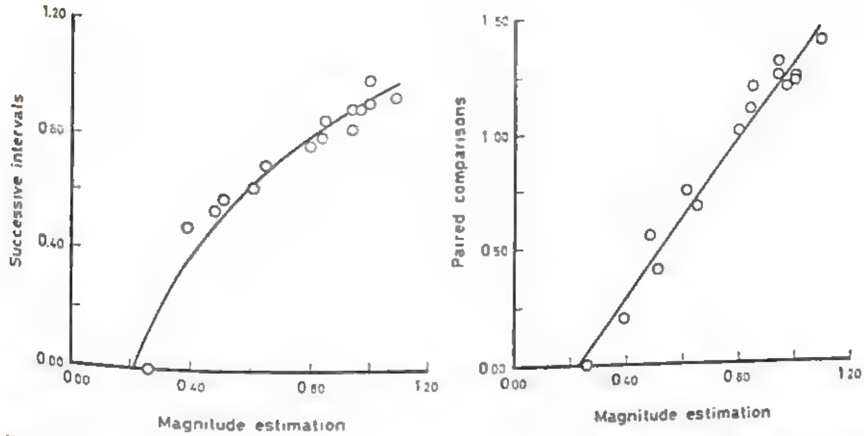


FIG. 3. TWO THURSTONIAN SCALES PLOTTED AGAINST MAGNITUDE ESTIMATION (For successive intervals, the function $y = a \log x + b$; for paired comparisons, a least-squares regression line)

SDs. One of the main purposes of this experiment was to study the structure of the dispersions of comparative judgments in relation to the hypothesized classification of similarities. Three classes of similarities have been mentioned: quantitiative, qualitative, and mixed. The expectation was that comparisons of the same type of similarity would show less dispersion than comparisons between different types of similarity. The dispersion terms resulting from scaling paired comparisons are reported in Table IV. The mean

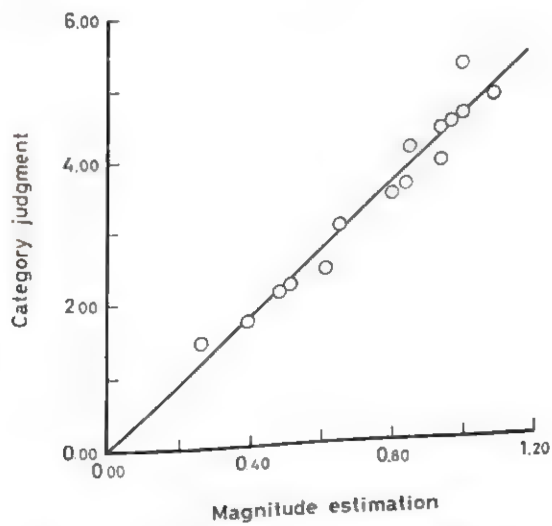


FIG. 4. ARITHMETIC MEANS OF CATEGORY JUDGMENTS PLOTTED AGAINST MAGNITUDE ESTIMATION AND A LEAST-SQUARES REGRESSION LINE

TABLE III
FACTOR ANALYSIS OF CATEGORY JUDGMENTS: ROTATED FACTOR LOADINGS

Stimulus	Factor I	Factor II	Factor III
AB	.30	−.32	−.35
AC	.37	.16	*.41*
AD	.20	.15	*.62*
AE	.22	*.79*	.03
AF	.24	*.40*	.37
BC	.09	−.02	*.66*
BD	.03	−.01	*.61*
BE	.06	*.70*	.26
BF	−.23	*.23*	.61
CD	.35	−.10	−.44
CE	*.60*	−.09	.19
CF	*.67*	.04	−.03
DE	*.57*	−.37	.07
DF	*.68*	.09	−.05
EF	.29	−.61	−.06

Note: Italicized values belong to stimuli used in identifying factors.

dispersion was 67.0 for within-type comparisons and 65.3 for between-type comparisons. The hypothesis was not confirmed.

However, the results of the factor analysis suggested another classification of similarities. As mentioned above, the three expected kinds of similarities did not emerge as separate factors. Instead, the three factors appeared to pertain to the pair of emotions being compared in a given pair of expressions. In accordance with this result, similarities were regrouped into three new clusters: depression-anger comparisons, happiness-anger comparisons, and happiness-depression comparisons. For each stimulus, the mean dispersion was computed both for comparisons within the same cluster

TABLE IV
*SD*s OF DIFFERENCES FROM PAIRED-COMPARISONS SCALING

	AB	AC	AD	AE	AF	BC	BD	BE	BF	CD	CE	CF	DE	DF	EF
AB		78	71	78	100	43	52	68	63	76	80	101	77	108	86
AC	78		41	59	48	39	46	84	66	63	62	77	56	77	80
AD	71	41		54	51	60	45	80	63	39	50	66	46	54	64
AE	78	59	54		51	74	64	52	62	63	56	69	84	72	89
AF	100	48	51	51		82	79	86	56	60	56	41	62	61	51
BC	43	39	60	74	82		39	66	48	71	62	100	60	76	64
BD	52	46	45	64	79	39		59	50	51	52	70	59	60	106
BE	68	84	80	52	86	66	59		60	79	86	99	89	87	99
BF	63	66	63	62	56	48	50	60		65	64	65	85	95	69
CD	76	63	39	63	60	71	51	79	65		47	58	63	53	56
CE	80	62	50	56	56	62	52	86	64	47		61	53	56	65
CF	101	77	66	69	41	100	70	99	65	58	61		64	44	62
DE	77	56	46	84	62	60	59	89	85	63	53	64		53	64
DF	108	77	54	72	61	76	60	87	95	53	56	44	53		57
EF	86	80	64	89	51	64	106	99	69	56	65	62	64	57	

Note: Decimal points preceding two-digit figures have been omitted.

and for comparisons with stimuli in the two other clusters. Within-emotion comparisons were excluded from this analysis. The means are reported in Table V, which shows that the new grouping gave the expected result. All stimuli except one had larger between-cluster than within-cluster dispersions.

The result may also be expressed as a point biserial correlation between size of dispersion and the dichotomy same cluster–different clusters. The correlation was −.40. Absolute size of difference between scale values also correlated with dispersions ($r = .12$). It was desirable to compute the partial correlation between dispersions and the dichotomy with size of difference held constant. The result could otherwise be explained by the somewhat smaller differences between scale values within clusters. The partial correlation dropped only to −.39, however.

The dispersion terms from magnitude estimation were also analyzed for information as to kind of similarity, using the classification suggested by the factor analysis. In Fig. 5 total *SD*s are plotted against arithmetic means. A corresponding plot for intraindividual *SD*s may be seen in Fig. 6.[17] The regression lines were fitted to points corresponding to between-cluster comparisons only, with the expectation that points corresponding to within-cluster comparisons would fall consistently below the regression line. All comparisons involving within-emotion stimuli were excluded from this analysis. As Figs. 5 and 6 show, the hypothesis of multidimensional struc-

TABLE V

MEAN *SD*s OF DIFFERENCES BETWEEN AND WITHIN CLUSTERS

Stimulus	Mean within cluster	Mean between clusters
AC	42	66
AD	49	58
AE	55	67
AF	64	60
BC	46	71
BD	43	62
BE	66	81
BF	59	67
CE	57	61
CF	56	73
DE	57	68
DF	51	73

Note: Decimal points preceding two-digit figures have been omitted.

[17] The intraindividual *SD*s were computed as the square root of the mean of the individual variances, the latter statistics having been computed for each *S* across three replications.

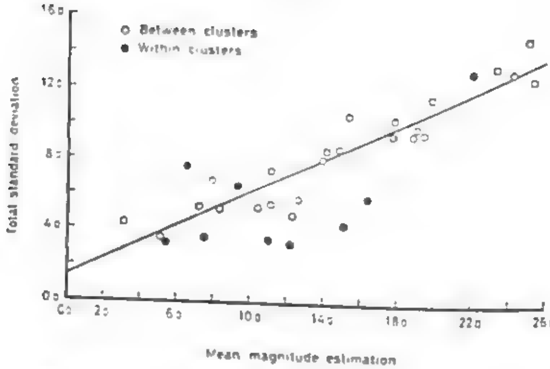


FIG. 5. TOTAL *SD*s PLOTTED AGAINST ARITHMETIC MEANS OF MAGNITUDE ESTIMATIONS
(Regression line fitted only to between-cluster points)

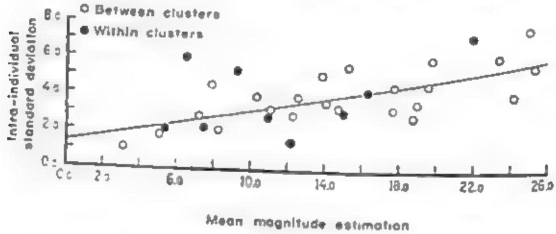


FIG. 6. INTRAINDIVIDUAL *SD*s PLOTTED AGAINST ARITHMETIC MEANS OF MAGNITUDE ESTIMATIONS
(Regression line fitted only to between-cluster points)

ture was not confirmed. A similar negative finding occurred in the earlier facial-expressions study.[18]

DISCUSSION

The high incidence of circular triads and the somewhat bad fit of scaling methods motivate a certain doubt as to the possibility of unidimensional measurement of similarity. However, there were also indications of rather large error variability in data, and for a more decisive check reliability should be taken into account.

On the other hand, relations among scaling methods were in complete agreement with previous studies, where the unidimensionality of the variable scaled can hardly be questioned. The paired-comparisons and magnitude scales were linearly related while the traditional successive-intervals scale was a nonlinear function of these two. In addition, a breakdown of the data by type of triad did not give evidence of more intransitivity of complex triads. This, and the fact that most triads had very few in-

[18] Sjöberg, *op. cit., J. exp. Psychol.*, 78, 1968, 429–435.

transitivities, suggests that *Ss* had very different kinds of multidimensionality, if any.

Furthermore, dispersion terms from paired comparisons proved to carry multidimensional information about the second-order similarity space which agreed with results from a factor analysis of category judgments of similarities. Dispersions of magnitude estimations did not give any clear evidence of multidimensional information.

The outcome of the factor analysis was difficult to interpret due to certain large loadings of the within-emotion pairs. However, its results did suggest a classification of similarities, which was clearly related to the dispersions of comparative judgments. This may be considered as a support for the tentative interpretation of the factor analysis.

Provided that the unidimensionality of the similarity variable can be trusted, the present results support several conclusions. Comparisons of scaling methods gave results in complete agreement with previous findings with variables characterizing single objects. This should give additional support to the generality of these relations. It is particularly interesting to note the linear relation between the paired-comparisons and magnitude scales. The curvilinear relation between traditional Thurstonian and magnitude scales has, so far, never occurred when the present method of paired comparisons was applied. This may be interpreted as indicating an error in traditional Case V assumptions; an error which has also been assumed by other authors.[19] Dispersion terms resulting from the use of paired comparisons did carry multidimensional information, as predicted by Thurstonian scaling theory. It may be argued that the occurrence of such meaningful structures, which agree with results from factor analysis, also throw doubt on the empirical validity of Case V (and Case IV) assumptions, since these include the assumption of zero or equal intercorrelations.

Thus, the present findings may be offered in support of the paired-comparisons scaling method used, and of Thurstonian scaling theory in general, but also as an argument against the traditional set of assumptions. The findings also support the possibility of using paired-comparisons data simultaneously for uni- and multidimensional scaling. Scale values are generated as in a usual

[19] G. Ekman and T. Künnapas, A further study of direct and indirect methods, *Scand. J. Psychol.*, 4, 1963, 77–80; L. V. Jones, Invariance of zero-point scaling over changes in stimulus context, *Psychol. Bull.*, 67, 1967, 153–164.

unidimensional method, but dispersion terms carry multidimensional information and may be processed as distance data. Finally, the findings did not support the initial hypothesis of quantitative, qualitative, and mixed similarities. Instead, both correlations of ratings of single stimuli and dispersions from paired comparisons indicated another trichotomy: happiness-anger, happiness-depression, and anger-depression. In a way, this is a more 'concrete' clustering of similarities than the one proposed initially. It is closely bound to the multidimensional structure of single pictures and puts an emphasis on variations in content rather than intensity of expressions.

Summary

This study investigated the dimensionality of similarity judgments and studied the relations among various methods for scaling similarity. The question of dimensionality was dealt with at two levels: (*a*) the basic one of whether a unidimensional scale can be constructed for a single individual on one occasion and (*b*) that of the structure inherent in inter- and intraindividual variability. Stimuli were pairs of facial expressions from the Lightfoot series, whose similarity was judged by 112 *S*s according to three response modes: paired comparisons (method of tetrads), category judgment, and magnitude estimation.

As for measurement properties, results were not entirely satisfactory, but there were no clear patterns of intransitivity, so the possibility of low reliability should be kept in mind. Relations among scaling methods were in agreement with previous results from scaling variables characterizing single objects. In particular, one may note a certain lack of proportionality among different magnitude scales, a nonlinear relation between the composite magnitude scale and a traditional successive-intervals scale, and a linear relation between magnitude estimation and paired comparisons. Consistent structure was found in intercorrelations from category-judgment data as well as in *SD*s of differences from paired comparisons—a finding which is in agreement with Thurstonian scaling theory and previous research. Stimuli (*i.e.* pairs of pictures) tended to cluster according to the pair of emotions involved.

A CONSIDERATION OF TWO ASSUMPTIONS UNDERLYING FIEDLER'S CONTINGENCY MODEL FOR PREDICTION OF LEADERSHIP EFFECTIVENESS

By Martin Fishbein, Eva Landy, and Grace Hatch,
University of Illinois, Urbana

In a recent series of papers, Fiedler has presented a model for the prediction of group performance.[1] As Fiedler points out, "the model is predicated on the assumption that the type of leadership behavior required for good group performance is contingent upon favorableness of the group-task situation for the leader."[2] More specifically, Fiedler identifies three major dimensions of the group-task situation. The *dimension of affective leader-group relations* refers to the personal relationship between the leader and key members of his group. According to Fiedler, this dimension reflects that which "is probably the most important single determinant of group processes which affect team performance. The liked and respected leader can obtain compliance from his group under circumstances which, in the case of a disliked or distrusted leader, would lead to open revolt."[3] The *task-structure dimension* refers to the clarity or ambiguity of the task. Here one may distinguish between highly structured, unambiguous tasks where the leader and his group members know exactly what needs to be done and the way to do it (*e.g.* a missile crew performing a countdown), and unstruc-

* Received for publication July 29, 1969. The study was supported by the Advanced Research Projects Agency, under ARPA Order 454, Contract NR 177-472, Nonr 1834(36), under which it appears as Technical Report No. 52. The authors express their indebtedness to Professors Fred E. Fiedler, Stanley M. Nealey, and Gordon O'Brien for their critical readings of an earlier draft of the paper.

[1] F. E. Fiedler, A contingency model of leadership effectiveness, in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, 1964, 149–190; F. E. Fiedler, The contingency model: A theory of leadership effectiveness, in H. Proshansky and B. Seidenberg (eds.), *Basic Studies in Social Psychology*, 1965, 538–555; F. E. Fiedler, *A Theory of Leadership Effectiveness*, 1967.

[2] F. E. Fiedler, A contingency model for the prediction of leadership effectiveness, Technical Report No. 10, Group Effectiveness Research Laboratory, University of Illinois, 1963, i. While most direct quotes were taken from this original manuscript, very similar statements are made in all the more readily available publications cited in n. 1.

[3] Fiedler, *op. cit.*, 1963, 4.

tured, ambiguous tasks where neither the leader nor the members can readily specify the manner in which such a task is to be executed (*e.g.* planning a program for a picnic). Fiedler views this task-structure dimension as the second most important determinant affecting team performance.[4] The *dimension of leader's position power* refers to the degree of formal or informal power inherent in the leadership position. A leader with high position power is one who can utilize rewards and sanctions, whose authority over his men is supported by the organization within which the group operates; a leader with low position power is one who is essentially restricted to using persuasion and other indirect means of influence.

By dichotomizing each of these dimensions, eight distinct types of group-task situations can be identified (Table I). Two points about these eight group-task situations should be noted. The first is that according to Fiedler, these situations vary along a continuum of favorableness to unfavorableness for the leader, *i.e.* differ with respect to the degree to which they permit the leader to "influence and control his group members."[5] The leader is seen as having maximum influence over, and control of, his group members in Group-Task Situation 1, and minimal influence and control in Situation 8. The second point is that each of these group-task situations may "require a different leader-group member interaction,"[6] *i.e.* may well 'demand' different types of leadership behavior in order for the group to operate at maximum efficiency. One type of situation may require a permissive, nondirective, considerate type of leader, while another group-task situation may require a controlling, managing, directive type of leader. Fiedler has argued that such is indeed the case.[7]

TABLE I

DIMENSIONS OF GROUP-TASK SITUATIONS

Group-task situation	Affective leader-member relations	Task structure	Leader's position power
1	Good	Structured	High
2	Good	Structured	Low
3	Good	Unstructured	High
4	Good	Unstructured	Low
5	Poor	Structured	High
6	Poor	Structured	Low
7	Poor	Unstructured	High
8	Poor	Unstructured	Low

[4] See n. 1.

[5] Fiedler, *op. cit.*, 1963, 13.

[6] Fiedler, *op. cit.* [in Berkowitz], 1964, 6.

[7] Fiedler, *op. cit.* [in Berkowitz], 1964.

One of the major variables investigated by Fiedler and his associates has been "the leader's esteem for his least preferred co-worker (LPC)."[8] There is a considerable amount of evidence indicating that High LPC leaders are more effective than Low LPC leaders in certain group-task situations, while Low LPC leaders are more effective than High LPC leaders in different types of group-task situations. In his earlier writings, Fiedler viewed the High LPC leader as a person who is permissive, nondirective, and considerate; and the Low LPC leader as a person who is controlling, managing, and directive.[9] More recently, however, he has emphasized a motivational, rather than behavioral, basis for distinguishing between (or describing) High and Low LPC leaders.[10] Fiedler now views the High LPC leader as a person who "obtains need satisfaction or reinforcement as a consequence of having experienced success in interpersonal relations"; the Low LPC leader as an individual who "obtains his need satisfaction or reinforcement through his achievement (or participation) in assigned group tasks."[11] According to this more recent approach, the High LPC leader tends to be more permissive and considerate than the Low LPC leader under normal conditions but not necessarily in all situations. In a given situation, the High LPC leader's motivation for achieving satisfying interpersonal relations may rather 'cause him to behave' in a more directive, controlling way than a Low LPC leader. Similarly, the Low LPC leader's motivational structure may 'cause him to act' in a permissive, considerate manner. Thus, although Fiedler still holds that High and Low LPC leaders behave differentially in given situations, he says that the specific types of behaviors they display vary as a function of their underlying motivational orientations.

Briefly then, the contingency model may be seen as an attempt to tie together, to integrate, the findings of 12 years of research which suggest that High and Low LPC leaders are differentially effective in different types of group-task situations. In particular, Fiedler hypothesizes that in situations that are very favorable or very unfavorable for the leader, a Low LPC leader is most effective; while in situations moderately favorable or moderately unfavorable

[8] See n. 1.

[9] *E.g.*, F. E. Fiedler, *Leader Attitudes and Group Effectiveness*, 1958.

[10] Fiedler, *op. cit.* [in Berkowitz], 1964; F. E. Fiedler, The effect of leadership and cultural heterogeneity on group performance: A test of the contingency model, *J. exp. soc. Psychol.*, 2, 1966, 237–264; Fiedler, *op. cit.*, 1967.

[11] Fiedler, *op. cit.*, 1967, 45.

for the leader, a High LPC leader is most effective. (Several validation studies have provided support for this as well as other hypotheses generated by the contingency model.)[12] His conceptualization assumes that High and Low LPC leaders are differentially effective in different group-task situations because (*a*) these situations call for different kinds of leadership behaviors for maximally efficient group performance and because (*b*) High and Low LPC leaders differ with respect to leadership styles and orientations, one type of leader (*e.g.* a High LPC leader) thus being better at meeting the demands of a particular given situation than another type of leader (*e.g.* a Low LPC leader). Further, it should be recalled that from Fiedler's point of view, these situational demands covary with the three dimensions of the contingency model, which dimensions are, he hypothesizes, of differential importance. That is, the kinds of leadership behaviors required for maximal group effectiveness vary most with the nature of the affective leader-member relations, next most with the task structure, and least with the leader's position power.

The major purpose of the present paper was to investigate these two basic assumptions underlying the contingency model: (1) that different group-task situations require different leader-group member interactions, *i.e.* demand different types of leadership behaviors, and (2) that these demands covary systematically with the three dimensions of the contingency model. In addition, an attempt was made to explore one possible reason for the differential behaviors of High and Low LPC leaders; for although Fiedler has presented evidence that High and Low LPC leaders do indeed behave differentially in given situations,[13] the determinants of this differential behavior have not been made explicit. While Fiedler has argued that the basis of this differential behavior is the different motivational orientations of High and Low LPC leaders, one may ask whether these motivational differences are directly reflected in behavior or whether they operate through an intervening variable. For example, High and Low LPC leaders may behave differently because they differently perceive the demands of a given situation and act in accordance with their perceptions, or because they similarly perceive the demands of a given situation but have character-

[12] Fiedler, *op. cit., J. exp. soc. Psychol.*, 2, 1966, 237–264.

[13] F. E. Fiedler, A review of research on AS_o and LPC scores as measures of leadership style, Technical Report No. 33, Group Effectiveness Research Laboratory, University of Illinois, 1966.

istically different modes of responding to it. The present study was thus also concerned (3) with the ways in which Ss who vary in LPC perceive the demands of group-task situations. Although the suggestion that Ss differing in LPC perceive the same demands seems more consistent with Fiedler's position, it should be noted that either alternative is consistent with the general notion that High and Low LPC leaders differ with respect to leadership styles and orientations. Further, it should be noted that these two alternatives are not mutually exclusive, and it is possible that both occur simultaneously.

Method

Procedure. The Ss were 141 undergraduates (men) who participated in the experiment as part of a course requirement. All Ss were tested simultaneously during a 60-min. session. Upon entering the experimental room (a large auditorium), Ss were randomly assigned to seats. After all were seated, questionnaire booklets were passed out. Each booklet contained the following: (1) *A standard form of Fiedler's LPC Scale.* This instrument consists of 25 eight-place bipolar adjective scales, in the Semantic Differential format. The Ss were asked to:

> Think of your least preferred coworker—that is, think of the one person you have had the most difficulty working with—and rate that person on the following scales. Remember, we are not necessarily asking you to think of the person you liked the least, but the one person you have had the most difficulty working with.

(2) *A handout describing the three dimensions of the contingency model,* with examples representing the end points of each dimension. More specifically, the handout provided verbal descriptions (or definitions) of 'good' and 'bad' affective leader-group relations, of highly 'structured' and highly 'unstructured' tasks, and of 'high' and 'low' power positions.[14] The handout was not attached to the booklet, and Ss could refer to it at all times. (3) *A series of Behavioral Description Questionnaires (BDQs),* which took the eight pages of the booklet following the LPC scale. The BDQ is based directly on Bales's method of interaction-process analysis.[15] Each questionnaire consisted of 12 items representing Bales's interaction-process categories; each item described a different kind of behavior (*e.g.* "he helps to clarify the situation by providing useful information"), and Ss were asked to indicate whether "the most effective leader" would engage in that behavior in a given group-task situation by checking an eight-place scale ranging from "very true" (8) to "very untrue" (1). The group-task situation itself was described at the top of the questionnaire (*e.g.* "Given a group-task situation where the leader has GOOD AFFECTIVE RELATIONS, where the group is working on a HIGHLY STRUCTURED TASK, and where the leader has a LOW POWER POSITION"), so that on each of the eight questionnaires the Ss were marking 12 eight-point

[14] Copies available from senior author.

[15] R. F. Bales, *Interaction Process Analysis,* 1951. Copies of the BDQ used in the present experiment are also available from the senior author.

TABLE II

ANALYSES OF VARIANCE: JUDGMENTS OF THE MOST EFFECTIVE LEADER'S TASK-ORIENTED BEHAVIOR

Source	df	Answer-giving behavior *SS*	*MS*	*F*	%T	%C	Question-asking behavior *SS*	*MS*	*F*	%T	%C
Between *Ss*	(140)	(6509.5)	(46.5)		(35.4)		(9851.0)	(70.4)		(34.7)	
LPC	2	13.8	6.9	<1	.1	1.0	124.2	62.1	<1	.4	5.2
Res. *Ss*	138	6495.7	47.1		35.4		9726.8	70.5		34.3	
Within *Ss*	(987)	(11853.8)	(12.0)		(64.6)		(18529.7)	(18.8)		(65.3)	
Variables (situations)	(7)	(1252.6)	(178.9)	(16.56)**	(6.8)	(88.3)	(2090.7)	(298.7)	(17.78)**	(7.4)	(87.6)
Affective relations (A)	1	718.1	718.1	38.20 **	3.9	50.6	1385.2	1385.2	49.12**	4.9	58.0
Task structure (B)	1	1.1	1.1	<1	.0	.1	16.8	16.8	<1	.0	.7
Position power (C)	1	410.0	410.0	32.54 **	2.2	28.9	485.4	485.4	19.04**	1.7	20.3
A × B	1	30.0	30.0	3.03	.2	2.1	25.1	25.1	1.96	.1	1.1
A × C	1	1.0	1.0	<1	.0	.1	170.1	170.1	13.10**	.6	7.1
B × C	1	71.4	71.4	9.52 *	.4	5.0	2.0	2.0	<1	.0	.1
A × B × C	1	21.0	21.0	3.50	.1	1.5	6.1	6.1	<1	.0	.3
Variables × LPC	(14)	(152.0)	(10.9)	(1.01)	(.7)	(10.7)	(172.3)	(12.3)	(<1)	(.6)	(7.2)
A × LPC	2	17.3	8.7	<1	.1	1.2	41.1	20.6	<1	.1	1.7
B × LPC	2	14.2	7.1	<1	.1	1.0	3.3	1.7	<1	.0	.1
C × LPC	2	64.0	32.0	2.54	.3	4.5	41.8	20.9	<1	.1	1.8
A × B × LPC	2	7.9	4.0	<1	.0	.6	25.9	13.0	1.02	.1	1.1
A × C × LPC	2	4.2	2.1	<1	.0	.3	28.8	14.4	1.11	.1	1.2
B × C × LPC	2	37.5	18.8	2.51	.2	2.6	20.8	10.4	<1	.0	.9
A × B × C × LPC	2	6.9	3.5	<1	.0	.5	10.6	5.3	<1	.0	.4
Variables × LPC × *Ss*	(966)	(10454.2)	(10.8)		(57.0)		(16266.7)	(16.8)		(57.3)	
A × Res. *Ss*	138	2599.4	18.8		14.2		3886.7	28.2		13.8	
B × Res. *Ss*	138	1442.5	10.5		7.9		2440.9	17.7		8.6	
C × Res. *Ss*	138	1738.4	12.6		9.5		3514.3	25.5		12.4	
A × B × Res. *Ss*	138	1363.3	9.9		7.4		1769.5	12.8		6.3	
A × C × Res. *Ss*	138	1453.4	10.5		7.9		1791.1	13.0		6.3	
B × C × Res. *Ss*	138	1035.7	7.5		5.6		1762.2	12.8		6.2	
A × B × C × Res. *Ss*	138	821.5	6.0		4.5		1102.0	8.0		3.9	
Total	1127	18363.3			100.0	100.0	28380.7			100.0	100.0

Note: %T and %C show percentage of total and controlled variance accounted for by each variable and its interactions.
* $p < .01$.
** $p < .001$.

TABLE III

ANALYSES OF VARIANCE: JUDGMENTS OF THE MOST EFFECTIVE LEADER'S INTERPERSONAL BEHAVIOR

		Positive socioemotional behavior					Negative socioemotional behavior				
Source	df	SS	MS	F	%T	%C	SS	MS	F	%T	%C
Between Ss	(140)	(6156.3)	(44.0)		(28.0)		(7747.6)	(55.3)		(36.6)	
LPC	2	71.0	35.5	<1	.3	1.8	10.6	5.3	<1	.1	.4
Res. Ss	138	6085.3	44.1		27.6		7737.0	56.1		36.6	
Within Ss	(987)	(15855.6)	(16.1)		(72.0)		(13409.6)	(13.6)		(63.4)	
Variables (situations)	(7)	(3725.5)	(532.2)	(42.91)**	(16.9)	(94.8)	(2072.1)	(296.0)	(25.74)**	(9.8)	(88.1)
Affective relations (A)	1	2487.3	2487.3	120.68**	11.3	63.3	1291.5	1291.5	71.35**	6.1	54.9
Task structure (B)	1	128.7	128.7	17.87**	.6	3.3	41.0	41.0	3.69	.2	1.7
Position power (C)	1	583.1	583.1	23.55**	2.6	14.8	378.0	378.0	24.54**	1.8	16.1
A × B	1	.0	.0	<1	.0	.0	9.8	9.8	1.22	.0	.4
A × C	1	502.6	502.6	50.26**	2.3	12.8	350.8	350.8	30.24**	1.7	14.9
B × C	1	6.1	6.1	<1	.0	.2	.0	.0	<1	.0	.0
A × B × C	1	17.7	17.7	2.53	.1	.4	1.0	1.0	<1	.0	.1
Variables × LPC	(14)	(134.1)	(9.6)	(<1)	(.6)	(3.4)	(269.4)	(19.2)	(1.67)	(1.2)	(11.5)
A × LPC	2	56.0	28.0	1.36	.3	1.4	46.1	23.1	1.28	.2	2.0
B × LPC	2	7.0	3.5	<1	.0	.2	40.4	20.2	1.82	.2	1.7
C × LPC	2	20.3	10.2	<1	.1	.5	122.2	61.1	3.97	.6	5.2
A × B × LPC	2	21.3	10.7	1.02	.1	.5	6.7	3.4	<1	.0	.3
A × C × LPC	2	16.2	8.1	<1	.1	.4	51.7	25.9	2.23	.2	2.2
B × C × LPC	2	7.8	3.9	<1	.0	.2	1.9	1.0	<1	.0	.1
A × B × C × LPC	2	5.5	2.8	<1	.0	.2	.4	.2	<1	.0	.0
Variables × LPC × Ss	(966)	(11996.0)	(12.4)		(54.5)		(11068.1)	(11.5)		(52.2)	
A × Res. Ss	138	2843.5	20.6		12.9		2495.2	18.1		11.8	
B × Res. Ss	138	994.1	7.2		4.5		1531.9	11.1		7.2	
C × Res. Ss	138	3417.4	24.8		15.5		2125.1	15.4		10.0	
A × B × Res. Ss	138	1445.2	10.5		6.6		1107.5	8.0		5.2	
A × C × Res. Ss	138	1386.7	10.0		6.3		1595.5	11.6		7.5	
B × C × Res. Ss	138	962.6	7.0		4.4		1204.6	8.7		5.7	
A × B × C × Res. Ss	138	946.5	6.9		4.3		1008.3	7.3		4.8	
Total	1127	22011.9			100.0	100.0	21157.2			100.0	100.0

Note: %T and %C show percentage of total and controlled variance accounted for by each variable and its interactions.
* $p < .01$.
** $p < .001$.

dimensions of the contingency model controlled 17.8% of the variance in beliefs about the most effective leader's positive socioemotional behavior, 11.1% of the variance in beliefs about his negative socioemotional behavior, 8.4% of the variance in beliefs about his question-asking behavior, and 7.6% of the variance in beliefs about his answer-giving behavior. Although this is clearly a significant and meaningful amount of variance to account for, it should be noted that between 80% and 90% of the variance in *S*s' judgments about the way the most effective leader should behave is essentially left unexplained. Since only 25% to 35% of this variance can be directly attributable to individual differences, this clearly implies that other variables relevant to the phenomena under consideration still have to be isolated.[19] Keeping this in mind, let us now turn to a consideration of the four major variables and their interactions.

LPC. Contrary to our third hypothesis, it can be seen that *S*s differing in LPC did not have differential expectations about the way the most effective leader should act in any of the eight group-task situations. Indeed, LPC and all its interactions with the three dimensions of the contingency model accounted for less than 1.5% of the total variance and less than 12% of the controlled variance in all four analyses. In no case did LPC or its interactions with other variables approach the .01 level of significance. This suggests that leaders who differ in their esteem for their least preferred coworker are *not* differentially effective in a given situation because they perceive the demands of that situation in different ways.

Group-task situations. The influence of the eight group-task situations as determinants of judgments about the way the most effective leader should behave is summarized on line 5 of Tables II and III. There it can be seen that the three situational variables and all their interactions accounted for approximately 90% of the controlled variance in all four analyses. Thus, almost all of the variance accounted for was directly attributable to variations in the group-task situations. Clearly then, as Fiedler has suggested, these different situations do seem to demand different types of leadership behaviors. However, in order to better understand the differences

[19] Although only 25%–35% of the variance can be directly attributed to individual differences, it should be noted that idiosyncratic use of the rating scales by *S*s also contributes to the variance. Indeed, there are many other sources of variance that can best be seen as 'noise in the system'; one would always expect some degree of uncontrolled variance.

between these situations, a consideration of each group-task situation variable and its interaction is necessary.

—*Leader-member affective relations.* In all four analyses, it can be seen that *Ss* expected the most effective leader to behave in a significantly different manner when he has good affective relations with his group members than when he has bad affective relations. More specifically, the most effective leader was expected to ask more questions, to give more answers, and to display more positive and less negative socioemotional behavior when he has good than when he has bad affective relations with his group members. The mean differences for each of the four behaviors may be seen in Table IV. In addition to influencing all four types of behavior, the leader-member affective relation was, as would be expected from Fiedler's model, the most important single determinant of expectations about the most effective leader's behavior (Tables II and III): in all cases, the affective-relations dimension accounted for more than 50% of the controlled variance in these expectations.

—*Task structure.* Contrary to expectations based on Fiedler's model, however, the degree of the task structure appeared to play a relatively minor role in determining expectations about the most effective leader's behavior. Even though task structure did have a significant effect on expectations about the most effective leader's positive socioemotional behavior (he was expected to show significantly more positive socioemotional behavior in structured than in unstructured situations; $\bar{X} = 17.51$ and 16.84, respectively, $p < .001$), it accounted for less than .6% of the total variance and less than 3.3% of the controlled variance in any of the four analyses. Here, however, it must be noted that the present analysis was concerned with *Ss'* expectations about a leader's behavior in hypothetical situations and *not* with actual behavior in real situations. Since

TABLE IV

Influence of Affective Leader-Member Relations on Expectations about the Most Effective Leader's Behavior

Type of behavior	Good affective relations	Bad affective relations	*F*
Positive socioemotional	18.66	15.69	120.68
Negative socioemotional	9.84	11.98	71.35
Answer-giving	18.90	17.30	38.20
Question-asking	18.45	16.24	49.12

Note: All *F*s significant beyond the .001 level of confidence. The higher the mean score, the more likely the most effective leader was judged to engage in the particular type of behavior. Scores could range from 3 to 24.

Fiedler's model is based on an analysis of these latter situations, the present finding does not necessarily mean that task structure is an irrelevant dimension in his analysis of ongoing groups, nor does it mean that leaders in a real situation do not take the task structure into account in determining their course of action. The finding does suggest, however, that the task-structure dimension may not be as important as Fiedler has indicated, and certainly deserves a closer and more critical analysis.[20]

—*Leader's position power.* As was the case with affective leader-member relations, *Ss* expected the most effective leader to behave in a significantly different manner when he has high position power than when he is in a position of low power. This effect was significant beyond the .001 level in all four analyses. The most effective leader was expected to ask fewer questions, to give more answers, and to show more negative and less positive socioemotional behavior when he has a high-power position than when he has a low-power position. Table V presents the mean differences for each of the four behaviors.

In addition, it should be noted that the leader's position power was the second most important determinant of expectations in all four analyses. As might have been expected, position power appeared to account for slightly more of the variance of task-oriented behaviors than of socioemotional behaviors. While position power

TABLE V

INFLUENCE OF LEADER'S POSITION POWER ON EXPECTATIONS ABOUT THE MOST EFFECTIVE LEADER'S BEHAVIOR

Type of behavior	High position power	Low position power	F
Positive socioemotional	16.46	17.73	23.55
Negative socioemotional	11.49	10.33	24.54
Answer-giving	18.70	17.49	32.54
Question-asking	16.69	18.00	19.04

Note: All *Fs* significant beyond the .001 level of confidence. The higher the mean score, the more likely the most effective leader was judged to engage in the particular type of behavior. Scores could range from 3 to 24.

[20] Reviewing an earlier draft of this paper, Gordon O'Brien suggested that task structure may have had a small effect in the present study because of the use of Bales's categories for describing leader behavior—categories primarily designed to handle behaviors in that one type of task situation in which a group attempts to solve an unstructured verbal problem. Although it is possible that significant differences might have been found had other behavioral categories been used, it should be recalled that Bales considers his categories capable of enveloping all behavior that can occur in any small face-to-face group. In keeping with this, Bales's interaction-process analysis has been used in many different contexts. Similarly, Fiedler and his associates have often used the BDQ in various group-task situations.

accounted for 28.9% of the controlled variance in expectations about answer-giving behavior and for 20.3% of the controlled variance associated with question-asking behavior, it accounted for only 16.1% of the variance of negative socioemotional behavior and 14.8% of the controlled variance of positive socioemotional behavior. Unlike affective relations, which appear to have a similar influence on all types of behavior, the leader's position power seems to be most important in influencing his task-oriented behavior and less important with respect to interpersonal behavior.

Affective Relations × Task Structure interaction. This interaction was not significant in any of the four analyses. Further, in no case did this interaction account for more than .2% of the total variance or more than 2.1% of the controlled variance. Thus the relative unimportance of the task-structure dimension as a determinant of expectations about the most effective leader's behavior was again demonstrated.

Affective Relations × Position Power interaction. In three of the four analyses, this interaction was significant beyond the .001 level of confidence. Like the power variable, this interaction appears to have differential influence on different types of behaviors. Although this interaction was the third largest determinant of variance in expectations in the three analyses where it was significant, it was considerably more important with respect to interpersonal behavior than with respect to task-oriented behavior. It accounted for 14.8% of the controlled variance in expectations about the most effective leader's positive socioemotional behavior and for 16.1% of the controlled variance associated with negative socioemotional behavior; but for only 7.1% of the controlled variance associated with question-asking behavior and for none of the variance associated with answer-giving behavior.

The means for the three significant interactions are shown in Table VI. In all three cases, it appears that when the leader is in a task situation having good affective relations, *Ss* did not expect the most effective leader to behave differently when he has high power than when he has low power. However, in those situations where the leader has bad affective relations with his members, the most effective leader was expected to behave quite differently depending upon his position power. More specifically, when there are poor affective relations and the leader has a high-power position, he was expected to ask fewer questions and to show more negative and

TABLE VI

INFLUENCE OF AFFECTIVE RELATIONS × POSITION POWER INTERACTION ON EXPECTATIONS ABOUT THE MOST EFFECTIVE LEADER'S BEHAVIOR

	Affective leader-member relations					
Position power	Good	Bad	Good	Bad	Good	Bad
High	18.61	14.30	9.86	13.11	18.18	15.19
Low	18.71	17.08	9.82	10.84	18.72	17.28
	Positive socioemotional behavior $F = 50.26$		Negative socioemotional behavior $F = 30.24$		Question-asking behavior $F = 13.10$	

Note: All Fs significant beyond the .001 level of confidence. The higher the mean score, the more likely the most effective leader was judged to engage in the particular type of behavior. Scores could range from 3 to 24.

less positive socioemotional behavior than when he has a low-power position.

Task Structure × Position Power interaction. Tables II and III show that this interaction reached the .01 level of significance in only one of the four analyses. Interestingly enough, this one significant interaction was with respect to the only type of behavior that was not influenced by the Affective Relations × Position Power interaction: namely, answer-giving behavior. Again, however, it must be noted that even though this one interaction was significant, it must be considered skeptically since it accounted for less than .5% of the total variance and for only 5% of the controlled variance of expectations about the most effective leader's answer-giving behavior. Further evidence of the relative unimportance of this interaction can be seen in Table VII, which presents the means for this interaction and shows that although the power variable had a large influence on expectations about question-answering behavior, task structure had relatively little. The interaction indicates that when a leader has a high-power position, he

TABLE VII

INFLUENCE OF TASK STRUCTURE × POSITION POWER INTERACTION ON EXPECTATIONS ABOUT THE MOST EFFECTIVE LEADER'S ANSWER-GIVING BEHAVIOR

	Task structure	
Position power	High	Low
High	18.98	18.42
Low	17.27	17.72

$F = 9.52,\ p < .01$

Note: The higher the score, the more likely the most effective leader was judged to engage in answer-giving behavior. Scores could range from 3 to 24.

was expected to answer slightly more questions when the task is structured than when it is unstructured. In contrast, when he has a low-power position, he was expected to answer slightly more questions when the task is unstructured than when it is structured. Neither of these differences, however, was significant at the .01 level of confidence.

Affective Relations × Task Structure × Position Power interaction. Tables II and III show that the triple interaction did not approach significance in any of the analyses. Further, and perhaps more importantly, in no case did this interaction account for more than .1% of the total variance or for more than 1.5% of the controlled variance.

To summarize then, although more than 90% of the controlled variance in expectations about the way the most effective leader should behave was attributable to variations in the group-task situation, almost all of this variance was accounted for by only two variables and their interaction. More specifically, Affective Relations, Position Power, and the Affective Relations × Position Power interaction accounted for 90.0% of the controlled variance of positive socioemotional scores, 85.9% of the controlled variance of negative socioemotional scores, 85.4% of the controlled variance of expectations about question-asking behavior, and 79.6% of the controlled variance of expectations about answer-giving behavior. Thus, while the data were consistent with Fiedler's notion that different group-task situations demand different kinds of leadership behavior for maximum group effectiveness, only two of the three dimensions isolated by Fiedler appear capable of consistently explaining a significant percentage of the variation in these demands. Further, it must be recalled that although these two dimensions did account for 80% to 90% of the controlled variance, they accounted for only 6% to 16% of the total variance. It is obvious that other variables associated with the group-task situation remain to be isolated. Indeed, it is worth noting that in his more recent articles, Fiedler has been doing just this. For example, he has considered the homogeneity-heterogeneity of the group and the amount of stress (either internal or external) under which the group is working. There can be little doubt that considerably more research is necessary before the various complexities of the group-task situation are isolated and the importance of these variables as determinants of 'demands' is assessed.

Before concluding, however, a note of caution must be reintroduced. Earlier in this paper, it was pointed out that while Fiedler's contingency model was based on an analysis of 'real' groups in 'real' situations, the present paper has been concerned with *Ss*' beliefs about the way 'the most effective leader' would behave in eight hypothetical situations. Clearly, although an *S* may expect the most effective leader to behave in the same manner on structured and unstructured tasks, this does not mean that the most effective leader will or does behave in the same manner. Similarly, although an *S* may expect the most effective leader to behave quite differently when he has high position power than when he has low position power, this does not mean the most effective leader will or does behave differently. To put this in a slightly different way, the difference between having high and low position power may appear quite large in the hypothetical situations presented in this paper, whereas in actual groups the distinction may have little, if any, practical significance.

Thus, in conclusion, although the present paper has presented evidence supporting Fiedler's assumption that different situations demand different leadership behaviors, evidence rejecting the hypothesis that *Ss* who differ in LPC perceive these demands differently, and some evidence questioning Fiedler's other assumption that the demands covary with the three dimensions of the contingency model in a specific manner, it must be recognized that these results can in no way be taken as a direct test of the model. Rather, they can only be viewed as supplementary to Fiedler's position. Where they agree with his position, they provide some convergent validity for his arguments; where they disagree, they merely raise questions about the validity of his assumptions and point to directions where further research with 'real' groups in 'real' situations must be conducted.

Summary

The present paper was an attempt to explore two of the basic assumptions underlying Fiedler's contingency model: (1) the assumption that different group-task situations "require a different leader-group member interaction," *i.e.* 'demand' different types of leadership behaviors; and (2) the assumption that these 'demands' covary systematically with the three dimensions of the group-task situation specified by the contingency model. The 141 undergraduates (men) who served as *Ss* rated the way they believed "the most

effective leader" should perform in each of eight group-task situations on a Behavioral Description Questionnaire. The results indicated that although these ratings of the most effective leader's behavior did vary across the different group-task situations, the ratings were significantly influenced by only two of the three group-task dimensions isolated by Fiedler: namely, the dimensions of leader-member relations and of leader's position power. Additional hypotheses related to the contingency model were investigated and discussed.

observed carefully. In the case of the concave face, the whole face shows apparent movement; the nose seems to swing round wildly as the observer moves, the eyes and cheeks move to a lesser extent, and the whole face appears to move in a 'rubbery' way.

In the experiments reported here, we investigated convex real faces and also flat faces, using a 'portrait' of a trained 'looker' (*L*) on a television screen. Concave faces were not investigated. The 'straight-at-*S*' gaze we used in the laboratory is a little different from that of real life. In the laboratory, *L* looked *S* straight in the bridge of the nose: in real life he would look him straight in the eye. It is immaterial, in real life, at which of *S*'s eyes the *L* looks, provided he look at only one of them: if *L* looks with his left eye at *S*'s right eye, and with his right eye at *S*'s left eye, both the *S* and *L* have the impression that *L* is looking straight through *S* (as indeed he is, since his eyes are converged at optical infinity). However, if *L* looks at only one of *S*'s eyes, *S* still perceives that *L* is looking straight at him. If *L* now switches his gaze to *S*'s other eye, then *S* perceives, paradoxically, that *L* has shifted his gaze and yet continues to look straight at him. (This 'eye-switching' is used as a technique in acting. Screen actresses are taught, when gazing into the hero's face in close-up, to switch their gaze alternately at each of his eyes in turn, as if 'searching his face.' They are not taught to gaze at the bridge of his nose.)

Experiment I reported here followed (and generally confirmed) the pioneering efforts of Gibson and Pick and of Cline.[1] Gibson and Pick required each *S* to judge whether or not he was being looked at by a trained 'looker' (*L*) 200 cm. distant from him. *L* fixated one of seven points, spaced at 10-cm. intervals along a horizontal line drawn on the wall behind *S*'s head, the middle point of the line having a position equivalent to the bridge of *S*'s nose. *L*'s head was fixed in one of three postures: straight ahead facing *S*, turned 30° to the left, and turned 30° to the right. For each of the three head postures, the frequency distribution of *S*'s 'yes' responses was plotted; *i.e.* judgments that '*L* is looking straight at me' were plotted. The means of the distribution were taken as an indication of *S*'s constant errors (CEs), and the standard deviations (*SD*s) as an indication of *S*s thresholds. Gibson and Pick found that thresholds were about the same when *L* looked askance at *S*, with head turned 30°,

[1] J. J. Gibson and A. D. Pick, Perception of another person's looking behavior, this JOURNAL, 76, 1963, 386–394; M. G. Cline, The perception of where a person is looking, this JOURNAL, 80, 1967, 41–50.

as when L gazed directly at S. Resolution of L's direction of gaze was approximately equal to Snellen acuity. The CEs, unlike thresholds, were altered by the head position: turning L's head caused an apparent shift of L's gaze. To paraphrase Gibson's results: when L aimed his head at S's left shoulder but looked askance at S's left ear, then S perceived L to be looking straight at him. Gibson called this an apparent shift *in the direction* of L's head movement. We propose to call it an apparent shift of L's gaze to S's right, *i.e. in the opposite direction* to L's head movement. The difference is purely verbal; Gibson used a null method, reporting the shift in L's gaze that was necessary for S to perceive L as looking straight at him. This null shift, as in all null methods, was in the opposite direction from the CE which was being nulled. We used a matching or reproduction method, and prefer to say that the direction of S's CEs were in the opposite direction to the head turn, both in Gibson's results and in our own. (The CEs that were obtained, below, in comparable conditions were, we might here note, about half as large as Gibson's.)

Cline presented S with a mirror image of L, 122 cm. away from him, who looked at a set of 13 targets scattered about his head, including one target located between his eyes. S was able to judge with a high degree of accuracy whether he was being looked at; the CE was a negligible .14 cm. to S's right, and the threshold (SD) was 1.55 cm., or a rotation of L's eye through 45 min. of arc. This represents a linear deviation of L's iris through only .18 mm., which from S's viewing distance represents an acuity of .51 min. of arc, somewhat better than Snellen acuity. This is also rather better than Gibson's figure of 1.1 min. of arc. Accuracy for judgments of gaze at off-center targets was considerably less than for the center target. In addition, irrespective of target position, there was an unexplained CE towards S's right, ranging in value up to 4.57 cm.

Cline then modified his equipment and replaced the target board with a horizontal row of five targets positioned at $-10°$, $-4°$, $0°$, $+4°$, $+10°$. The L's head could assume different positions: (*a*) always straight ahead directly facing S, (*b*) always turned 30° to the right, and (*c*) turned on each trial in the direction of gaze, so that L's eyes were always centered in their sockets. With L's head turned 30° to the right, the threshold value for the center target rose considerably, from 1.55 cm. to 5.29 cm., a finding at variance with that of Gibson and Pick, who reported no such loss of acuity. The acuity for targets either side of the center was also decreased,

though not to the same extent. Large CEs were reported "*in the direction* of the turn of the head." Cline did not make it absolutely clear what this means, but he remarked that his results agreed with Gibson's. So it appears that he found, as did Gibson and the present writers, that if *L*'s head pointed toward *S*'s left shoulder, then *L*'s gaze at *S*'s left ear seemed to *S* to be directed at *S*'s nose. (We prefer to call this a CE "in the direction opposite to the head turn.") When *L* changed his head posture for each line of regard, keeping his eyes stationary relative to his head, then all target positions had the same threshold (4° to 6.75°), except for the center target, whose threshold was lower (2.8°). CEs were small and again "in the direction of turn of eyes and head."

The present experiments extended the analyses made by Gibson and Pick and by Cline, of the stimulus for the perception of looking behavior in another person. Gibson and Pick studied a single case of the perception of where a person is looking: the perception of being looked at. Cline extended this by measuring CEs and thresholds for a wider range of gaze directions. We have varied some other stimulus parameters and have measured the linear function which relates the actual to the perceived direction of gaze. Gibson and Pick discovered the shift of the PSE when *L*'s head was turned, *i.e.* the head-turn effect. In the experiments below we discovered also that *S*s consistently overestimate all deviations of *L*'s gaze from the straight ahead, *i.e.* show the overestimation effect. We have tried to relate our findings, and those of Gibson and Pick and of Cline, to the physical stimulus provided by *L*'s eyeball. We have also investigated the possibility that the effects obtained with a human *L* could be reproduced by manipulating the geometry of the artificial eyeball. We present the hypothesis that the main determinant of judged direction of gaze is the centering of the pupil in the visible part of the eye.

Experiment I

Method. Two procedures were used. In the first (No TV), *S* looked directly at *L*'s head. In the second, *S* looked at a TV picture of *L*'s head. In each procedure, *S*'s task was to judge the direction in which *L* was looking. *Procedure 1.* The *L* sat facing *S*, at a distance of 84 cm. (Fig. 1). They both used chin rests. Midway between them was a horizontal scale symmetrically placed normal to the line of sight, and 6 cm. above eye level. The scale was calibrated in degrees of visual angle on one face, so the markings were visible to *L* but invisible to *S*. On either side of the center, nine target points were arranged at the values −20°, −15°, −10°, −5°, 0°, +5°, +10°, +15°, +20°, which *L* fixated.

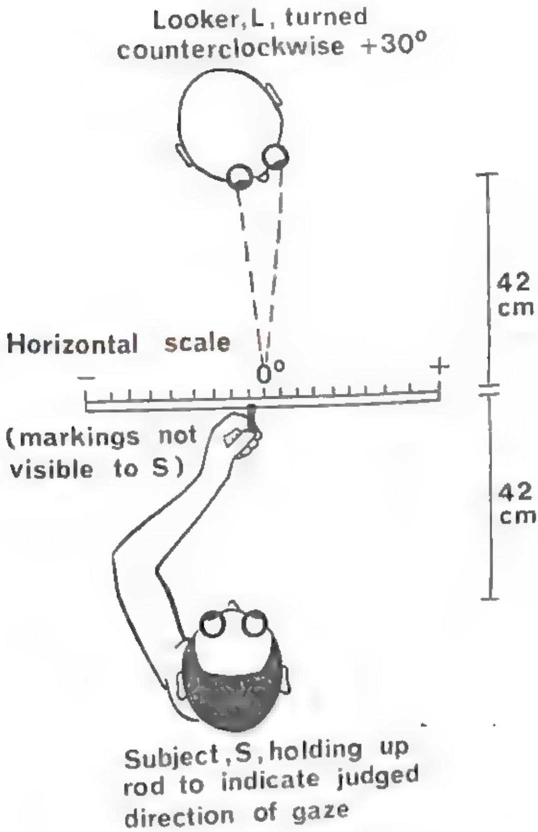


Fig. 1. Procedure 1, Experiment I

Negative values refer to targets on *S*'s left, positive values to targets on *S*'s right. On each trial, *S* held a short rod up to the scale to mark the point at which he thought *L* was looking, and *E* recorded the reading. *L* was a male with small green eyes. Five practice trials were run, then *L* fixated the nine targets in random order until each target point had been fixated three times. There were three conditions, distinguished by the posture of *L*'s head: straight ahead facing *S* (Normal to Subject); turned clockwise to −30°, to *S*'s left; and turned counterclockwise to +30°, to *S*'s right. In each condition a different random order of fixations was used.

—*Procedure 2.* This was essentially the same, but now *S* viewed *L*'s face displayed on a TV screen by means of a closed-circuit TV link (Fig. 2). A Pye Lynx TV camera was placed where *S* had been sitting, and *S* now sat on another chair, facing a 17-in. TV monitor at a distance of 84 cm. *L* still fixated targets on the original scale, but midway between *S* and the TV screen was placed an identical scale, on which *S* indicated the points at which he judged the face on the screen to be looking. The distance from *S* to the center of the TV screen was the same as from *L* to the camera, namely 84 cm. At this distance the TV picture of *L*'s face was exactly life-size. The original three conditions of *L*'s head posture were now combined with three positions of the TV

screen giving nine conditions in all. The positions of the TV screen were: straight ahead facing S (TV Normal), turned clockwise $-30°$ to S's left (TV 30°C), and turned counterclockwise $+30°$ to S's right (TV 30°CC). In all other respects Procedure 2 was the same as Procedure 1. Each S made a total of 324 judgments in the two procedures combined.

Subjects. The Ss were six men graduate students in psychology, 22–33 yrs. old. All had normal or corrected vision. They were instructed to indicate where L was apparently looking, as described above, but they were naïve as to the purpose or expected outcome of the experiment.

Results. The results are plotted in Fig. 3 (mean of six Ss). A regression line was fitted to the data for each condition, with the aid of an Olivetti Programma 101 computer. The regression equations are printed next to each curve in Fig. 3 and take the form

$$y = mx + c,$$

where y is S's judgment of where L was looking, x is the point at which L was actually looking, c is the intercept of the regression curve, *i.e.* the CE made by S, negative values to his left, positive values to his right, and m is the slope of the regression curve, the

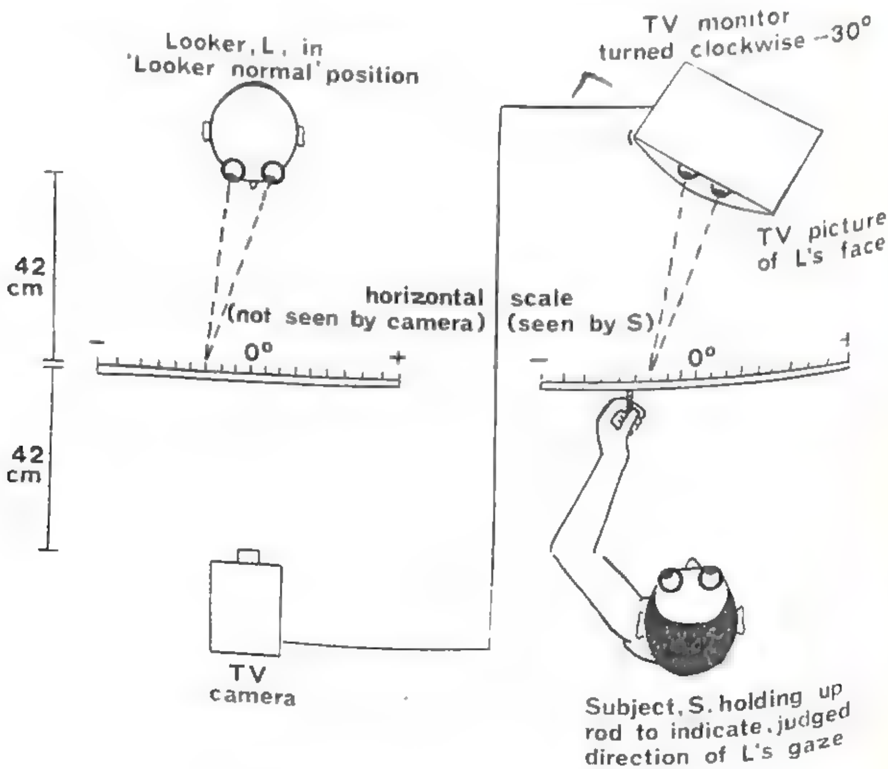


FIG. 2. PROCEDURE 2, EXPERIMENT I

factor or systematic error by which S over- or underestimated the angle of L's gaze. This m would be less than 1 if S underestimated and greater than 1 if he overestimated. In practice, it was found that S always overestimated the deviation of L's gaze from the straight-ahead position, and m was always greater than 1. For example, when L's head was turned $-30°$ clockwise in the No-TV arrangement of Procedure 1 (Fig. 3), it was found that: $y = 1.65x + 3.85$. This shows that the direction of L's gaze, expressed as a deviation of his eyes from the 'straight-at-S' position, was overestimated by 65% ($m = 1.65$), and superimposed on this systematic error of overestimation was a constant shift of S's judgments by +3.85 to S's right ($c = 3.85$).

—*All conditions: Systematic error of overestimation: Parameter m.* In all conditions S overestimated the deviation of L's gaze along the horizontal axis, by amounts ranging from 50% to 87% (m = 1.50 to 1.87). This means, roughly, that if L gazed at S's ear, it seemed to S that he was gazing at S's shoulder. Since this overestimation was a percentage error, the absolute error in degrees was least for gazes aimed straight at S and greatest for gazes aimed well away from him. The percentage errors of overestimation were least in the simplest conditions, which were also those which S reported as being easiest, namely when no TV link was used and when L's head was in the straight-ahead posture. The condition in which least overestimation occurred (50%) combined both these factors. As conditions were made more complex (*i.e.* TV introduced, TV screen turned, L's head turned), so percentage overestimations increased, the largest single factor being L's head posture. This overestimation effect appears to be newly discovered.

—*L's head posture: CEs: Parameter c.* When L's head was turned in one direction, his gaze seemed to S to deviate in the opposite direction. When L's head was turned clockwise to $-30°$ (to S's left), S perceived L's gazes as apparently shifted by an average CE of $+7.7°$ (to S's right) when results were averaged over all TV conditions. Similarly, when L's head was turned counterclockwise to $+30°$ (to S's right), his gazes were apparently shifted by an average of $-5.44°$ (to S's left). This confirms the reports of Gibson and Pick and of Cline.

—*TV-screen positions: CEs: Parameter c.* When the TV screen was turned, L's gaze appeared to S to deviate in the *same* direction. When the screen was turned clockwise to $-30°$, L's gazes were apparently shifted by an average CE of $-1.81°$ (results averaged

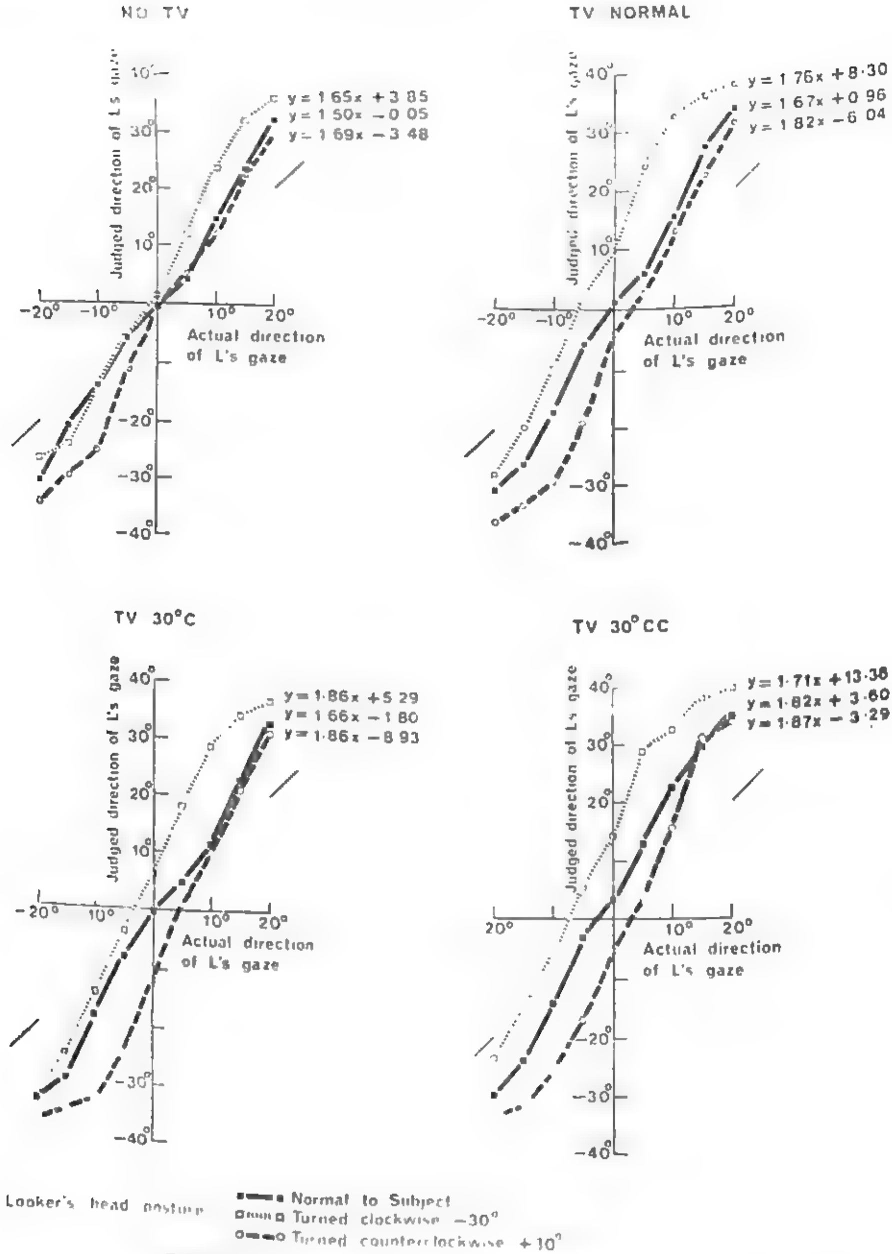


FIG. 3. *Ss*' JUDGMENTS OF *L*'S DIRECTION OF GAZE: EXPERIMENT I

over all head postures). Similarly, when the TV screen was turned counterclockwise to +30°, *L*'s gazes were apparently shifted by an average CE of +4.56°.

—*L's head posture vs. TV position: Comparison of CEs: Pa-*

rameter c. The CEs produced by turning *L*'s head were about twice as great as the CEs produced by turning the TV screen, and they were of *opposite* sign. Turning *L*'s head from −30° to +30° shifted the CE from +7.7° to −5.44°, a shift of −13.14° or −21.9%, whereas turning the TV through the same angle shifted the CE from −1.81° to +4.56°, a shift of +6.37° or +10.6%.

EXPERIMENT II

In this experiment we investigated judgments of vertical deviations of gaze. Conditions were the same as for Procedure 1 (No TV) of Experiment I. *L* again sat facing *S*, at a distance of 84 cm. The scale bearing the targets was now placed vertically, running from +30° (upward) through 0° (level, straight ahead) to −20° (downward). (The scale could not be extended further downward because *L*'s lowered lids hid his eyes from *S*.) The scale was placed midway between *L* and *S* and 6 cm. to the right of *S*'s line of sight. The *S*s were six graduate students in psychology (five men, one woman), 22–33 yrs. old. Three of the men had also served in Experiment I.

Results. The results are plotted in Fig. 4 (mean of six *S*s). They are rather uninteresting, since they show that *S* perceived *L*'s directions of gaze more or less veridically. The great discrepancy in results between judgments of horizontal and of vertical gazes no doubt reflects the considerable difference in the stimuli provided to *S* by the *L*'s eyeball. When the eyes move sideways, the amount of white (sclera) visible on either side of the pupil varies approximately according to a sine law. But when the eyes move vertically, the amount of sclera visible varies in a nonlinear manner. When the eyes look upward, sclera becomes visible below the pupil but not above, and when the eyes look downward, no sclera is visible above or below the pupil. This is because of the different behavior of the upper and lower eyelids. The upper lid tracks the upper edge of the pupil over the whole range of its vertical movements up and down, operated by an efficient position servo, whereas the lower lid moves downward for downward gazes but does not move upward for gazes above eye level. No trace of this nonlinearity was noticed in our results.

EXPERIMENT I AND II: DISCUSSION

Experiments I and II yielded three results which require explanation. They were: the TV-screen-turn effect, an apparent displace-

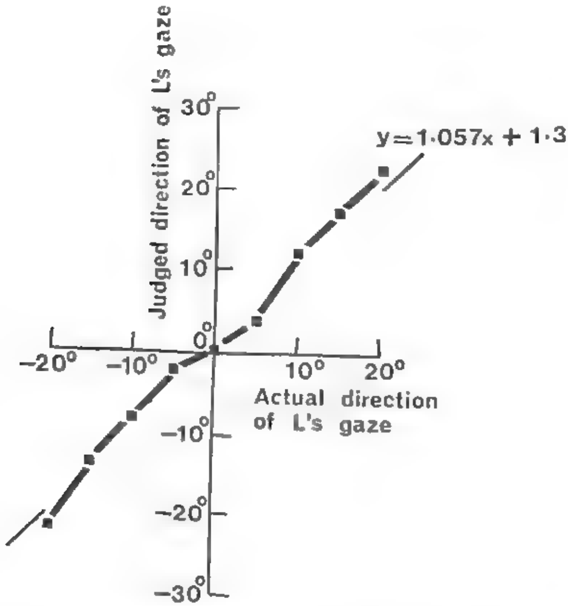


FIG. 4. *Ss'* JUDGMENTS OF *L*'S DIRECTION OF GAZE: EXPERIMENT II

ment of the perceived direction of *L*'s gaze, in the same direction as the turn of the TV screen; the head-turn effect, an apparent displacement of the perceived direction of *L*'s gaze, away from the direction of *L*'s head turn; and the overestimation of the deviation of *L*'s gaze from the straight ahead. We here put forward some hypotheses which attribute the TV-screen-turn effect to the curvature of the TV screen; and the head-turn effect and the overestimation to the special geometry of the eyeball.

The *TV-screen-turn effect* probably arose because of the convex curvature of the TV screen. All TV screens are convex, with the exception of projection TVs, which are now quite rare and which we were unable to obtain. The radius of curvature of the TV screen we used was approximately 30 cm. so, strictly speaking, our TV image was not a 'flat face' but a convex face, albeit less convex than a real face (of flesh and blood). Any such convexity made the TV image a little less like a flat painting, and a little more like a three-dimensional statue. When a statue is rotated, its line of sight rotates with it, as would a rod sticking rigidly straight out of its eye. The same is true, to a lesser extent, of a convex TV image. The greater the convexity, the more the direction of gaze is likely to rotate per degree of rotation of the portrait. This effect can be verified using any pic-

ture postcard of a portrait looking straight at one. Tilt the card: the portrait continues to look unwaveringly at you. Now curve the card so that it is convex: the direction of gaze now moves in the direction of tilt, and the greater the convexity, the greater this effect. Now curve the card the other way, so that it is concave, and the effect is reversed: the direction of gaze now moves in the direction opposite to the tilt, and the greater the concavity, the greater this effect. We regard this effect as of only minor interest.

The *overestimation* effect is more interesting, and it appears to be a newly discovered effect. It occurs, in our opinion, because only *part* of the front surface of the (spherical) eyeball is visible through the aperture of the eye socket. So, for example, if the pupil moves through one-tenth of the eyeball's circumference, it will move through more than one-tenth of the eye's visible surface. *S*'s judgments are made on the basis of the visible surface of the eye, not on its hidden parts. Therefore any eye movement is judged as larger than it really is.

The *head-turn effect* occurs for similar reasons. When the head turns, the eye sockets turn with it. If the head and eye sockets turn while the eyes are kept fixed on a stationary point, the position of the iris changes relative to the socket and eyelids around it. It is this change which alters the judged position of the eye, for the position of the eyes is judged largely by the position of the iris relative to the eyelids. The head-turn effect is independent of *S*'s judgments of the position of *L*'s head, or of any gestalt-like qualities of the face, since the effect still occurs when *L* looks through a peephole in a screen which hides his head from *S*. As *L* turns his head behind the peephole while keeping his gaze fixed on a stationary point, *S* perceives *L*'s gaze as moving in the opposite direction to the head turn.

EXPERIMENT III

Method. The same apparatus and procedure were used as before, but the human *L* was replaced by an artificial eye, the direction of gaze of which *S* judged. This 'eye' consisted of a table-tennis ball 37 mm. in diameter, surrounded with an artificial socket: a hole of 27 mm. in a black iris diaphragm. Thus the diameter of the socket was ¾ that of the 'eyeball.' The 'pupil' was actually a hole, as it is in the human eye, formed by the end of a 6-mm. diameter tube, which ended flush with the front of the ball and extended out through the back. The artificial eye was pivoted about a vertical axis passing through its center by means of an offset crank. There was thus no vertical spindle passing through the ball, and the crank was concealed from view. Both the socket and 'eye' pivoted about the same axis; thus the socket could be rotated while still remaining closely fitted to the 'eyeball.' The 'eye'

was aimed at target points positioned from —30° to +30° along a scale, situated midway (42 cm.) between it and the *S*. as before. *E* aimed the 'eye' by sighting through the tube. Each *S* judged the direction of gaze of the artificial eye while sitting, in turn, in three positions arranged in a 60° arc around the 'eye': a central position directly facing the 'eye' (Normal), a position 40 cm. to *S*'s left of the center position (Clockwise —30°), and a position 40 cm. to *S*'s right of the center position (Counterclockwise +30°). Thus the 'eye' was always set to gaze at the same range of target positions, but *S* changed his position for each of the three conditions.

Subjects. The *S*s were six graduate students in psychology (four men, two women), 22-33 yrs. old. Three of the men had also served in Experiments I and II.

Results. The results are plotted in Fig. 5 (mean of six *S*s). They show that all deviations of gaze of the artificial eye from the straight-ahead (0°, Normal) target were overestimated by about 63%. This parallels the results of Experiment I for real, human eyes. The results were almost identical for the three conditions: the position in which *S* was sitting made virtually no difference in his judgments of direction of gaze.

Experiment IV

The procedure was the same as in Experiment III except that each *S* now made his judgments of the direction of gaze while al-

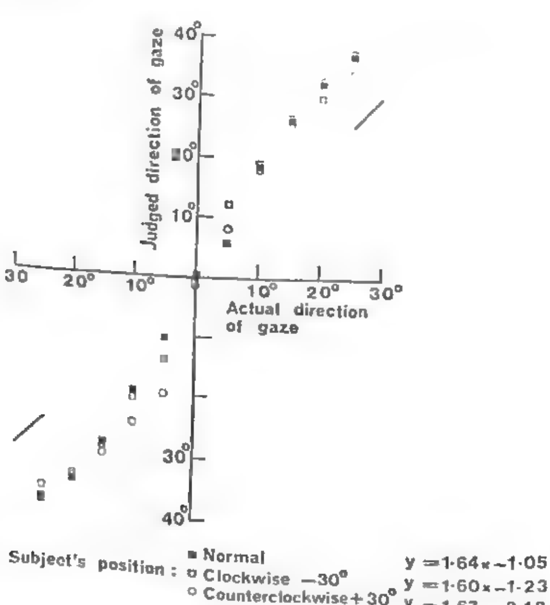


FIG. 5. *S*s' JUDGMENTS OF ARTIFICIAL EYEBALL'S DIRECTION OF GAZE: EXPERIMENT III

ways sitting in one position, directly facing the artificial eyeball. The independent variable was now the position of the diaphragm which formed the artificial eyeball's socket. The four conditions were: with the diaphragm wide open (50 mm.; termed No Diaphragm), a control condition in which the whole ball (37 mm. in diameter) was visible; with the diaphragm stopped down to 27 mm., *i.e.* to ¾ of the ball's diameter, placed normal to *S*, thus simulating the *L*'s straight-ahead (Normal) condition of Experiment I; with the diaphragm (27 mm.) rotated −30° clockwise, simulating the '*L*'s head rotated −30° clockwise' condition of Experiment I; and with the diaphragm rotated +30° counterclockwise, simulating the '*L*'s head rotated +30° counterclockwise' condition of Experiment I. The order of conditions was assigned at random to each of the six *S*s.

Results. The results are plotted in Fig. 6. With the diaphragm wide open, deviations of gaze were overestimated by 28% ($y = 1.28x + 0.14°$). It appears that deviations of gaze are somewhat overestimated, even when there is no diaphragm occluding part of the spherical surface of the eyeball. With the diaphragm normal, overestimations of deviations of gaze rose to 56% ($y = 1.56x - 0.12°$), twice that found with the diaphragm wide open. This parallels the 50% overestimation of *L*'s gaze found in Experiment I. With the diaphragm turned −30° clockwise, all judgments were shifted by 15° counterclockwise, *against* the direction of socket rotation ($y = 1.28x + 15.4°$). This parallels the shift of judged gaze *against* the direction of *L*'s head turn in Experiment I. In addition, overestimations of gaze deviations fell to 28%; we do not know why. With the diaphragm turned +30° counterclockwise, all judgments were shifted by 16° clockwise, against the direction of socket rotation ($y = 1.07x - 16°$). Once again this parallels the results of Experiment I. In addition, overestimations of gaze deviations fell to 7%; again we do not know why. In sum, *S*s overestimated the deviation of gaze of the artificial eye from the straight ahead in all conditions. This effect was increased by reducing the size of the artificial socket. Turning the diaphragm (*i.e.* rotating the socket) in one direction caused all judgments of gaze to be shifted in the opposite direction.

Experiment III and IV: Discussion

These two experiments showed that the overestimation effect and the head-turn effect can easily be generated by manipulating the

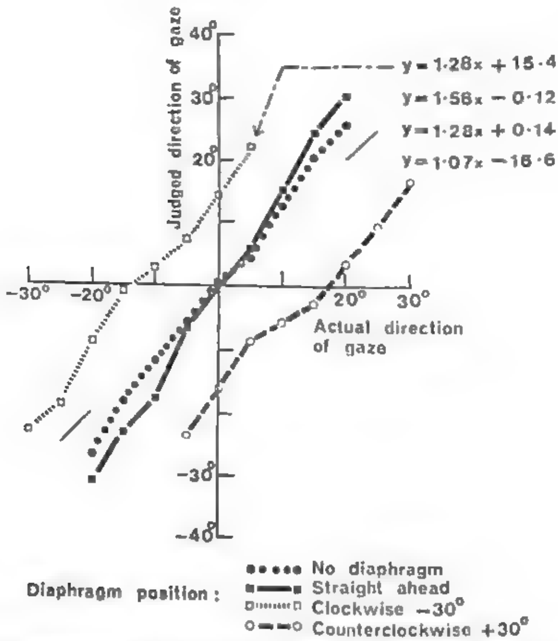


FIG. 6. *S*s' JUDGMENTS OF ARTIFICIAL EYEBALL'S DIRECTION OF GAZE: EXPERIMENT IV

geometry of an artificial eyeball. They confirm our hypothesis that the main determinant of judged direction of gaze is the centering of the pupil in the visible part of the eye. We could alter the apparent direction of gaze either by moving the pupil, or by exposing different amounts of white on either side of a stationary pupil. We measured the position of the pupil in this socket for various directions of gaze, both for *L*'s eye and for the artificial eye, by viewing the eyeball through a telescope with a graduated scale. We found that *S*'s judgments of *L*'s gaze, including the head-turn effect and the overestimation effect, could be predicted in a general way from the physical centering of the real or artificial pupil in its socket. Thus, *S* was responding largely to the physical stimulus of *L*'s eyeball. However, there was a small discrepancy between the artificial eyeball stimulus and *S*'s judgments, and a larger one between *L*'s eyeball stimulus and *S*'s judgments.

The centering of the pupil is not the only determinant of judged direction of gaze. We found that the apparent gaze of the artificial eyeball could be displaced to the left by tilting the diaphragm surrounding it to the right, even when the diaphragm was set at 50 mm. in diameter—larger than the eyeball and not occluding any part of it. This seems to be a kind of 'tilt contrast' effect. The effect was

small, but quite clear. In addition, as Luckiesh has pointed out, a pair of eyes in a portrait will tend to look in the same direction as the rest of the face.[2] So although the positioning of the pupils is a sufficient condition for the perception of a particular direction of gaze, it is not a necessary one. Other factors play a part in the complex processes that occur when one tries to judge where somebody is looking.

Summary

In Experiment I, six *S*s judged the direction of gaze of a trained 'looker' (*L*), either with *L* actually across from him or on a closed-circuit TV link. *L* fixated target points arranged at −20°, −15°, −10°, −5°, 0°, +5°, +10°, +15°, and +20° along a horizontal scale. When fixating the target at 0°, *L* was gazing straight at *S*. Linear regression functions were fitted to the data. It was found that *S* consistently overestimated, by 50% to 87%, the angle by which *L*'s gaze was turned away from him. Turning the TV screen 30° to left or right resulted in a 10.6% constant error (CE; here demonstrating an apparent shift in *L*'s direction of gaze) in the same direction as the TV screen was turned. This was probably due to the curvature of the TV screen. Turning *L*'s head 30° to the left or right resulted in a CE about twice as large (21.9%), and in the direction opposite to the turn of the head. The greatest CEs occurred when the head was turned in one direction and the TV screen in the other. In Experiment II, *S*s were able to judge *L*'s directions of gaze along a vertical axis almost veridically.

In Experiments III and IV, with the same setup, six *S*s judged the direction of gaze of an artificial eye along a horizontal scale. Results paralleled those of Experiment I. *S*s again consistently overestimated deviations of gaze. Reducing the size of the artificial socket increased the overestimation in the predicted direction, from 28% to 56%. Turning the artificial socket (equivalent to turning *L*'s head) resulted in a CE of 50.5% in the direction opposite to the turn, as predicted. *S*'s viewing position had virtually no effect on his judgments. It was concluded that judgments of direction of gaze are determined principally by the position of the pupil in the visible part of the eye.

[2] M. Luckiesh, *Visual Illusions*, 1922.

COGNITIVE DEPENDENCE ON ADDITIVE AND CONFIGURAL CUE-CRITERION RELATIONS

By BERNDT BREHMER, University of Umeå, Sweden

In the last few years there has been an increase in studies on how humans integrate information from various sources into judgments. In part, this research has grown out of the controversy over the relative efficiency of clinical and statistical prediction.[1] Following the suggestion by Meehl, studies in this area are now investigating "the cognitive activity of the clinician"[2] instead of continuing to compare the accuracy of clinicians with that of regression equations. That is, the emphasis has changed from a comparison of the relative *accuracy* of the two methods to a comparison of how clinicians and regression equations use the available information to *arrive at* predictions. This is, potentially, a more fruitful approach to the problem, since if there are any differences in accuracy between clinical and statistical prediction, they must stem from differences between how clinicians and regression equations utilize the information available to them. Once these differences are known, it should be possible to derive the relative efficiency of clinical and statistical prediction for all possible kinds of prediction tasks, and further comparison of the accuracy of the two methods would not be necessary.

The typical procedure in studies of the process of clinical judgment has been to present a judgment task (*e.g.* a series of MMPI profiles) to the *S*s (*e.g.* clinical psychologists, physicians, or college sophomores) and ask for a series of judgments (*e.g.* whether the patient depicted in a profile was neurotic or psychotic). After the judgments have been collected, the relations between these judgments and the various sources of information have been analyzed

* Received for publication June 30, 1969. The study was supported in part by a grant from the Swedish Council for Social Science Research and in part by NSF Grant GS-1699 to the Institute of Behavioral Science, University of Colorado, Boulder, Colorado. It is Publication No. 127 of the Institute. The paper was written when the author spent the academic year 1968–1969 at the Institute as a visiting scientist sponsored by the Sweden-America Foundation.

[1] P. E. Meehl, *Clinical vs. Statistical Prediction,* 1954.

[2] P. E. Meehl, The cognitive activity of the clinician, *Amer. Psychol.,* 15, 1960, 19–27.

by means of some standard technique such as multiple regression statistics or analysis of variance.

The typical finding in these studies has been that most of the variance in the response systems of the *Ss* can be accounted for by means of a linear multiple-regression equation; *i.e.* the clinicians seem to utilize the information in the same linear, additive fashion as the actuary.[3] As Hoffman pointed out, these results are at variance with what the *Ss* themselves say that they are doing.[4] If, for instance, a clinical psychologist is asked how he uses his information, he will often say that he uses it configurally, *i.e.* that his interpretation of one kind of information is contingent on what other sources of information show about his patient. Still, the majority of the studies indicate that most or all of the variance in his response system can be accounted for by a linear multiple-regression equation. When evidence for configurality has been found,[5] the configurality has accounted for a miniscule proportion of the variance; furthermore, it has not been entirely clear whether the configurality is *true* configurality or is due to inconsistencies in the *Ss'* use of the response scales.[6] Thus, there is at present very little evidence that humans use information in any other way than the linear, additive (or averaging) one. These results raise the interesting question of whether this form of information utilization is inherent in the human information-processing system or whether there are alternative explanations of the evidence.

Research on human judgment seems to have proceeded from the explicit or implicit hypothesis that there is, somewhere in *S*, a *single* process of judgment, the characteristics of which can be extracted from an analysis of the relations between the information available to the judge at the time of judgment and his responses. Thus, Anderson pursues the question of additive vs. averaging models of the

[3] L. R. Goldberg, Simple models or simple processes? Some research on clinical judgments, *Amer. Psychol.*, 23, 1968, 483–496; K. R. Hammond and D. A. Summers, Cognitive dependence on linear and nonlinear cues, *Psychol. Rev.*, 72, 1965, 215–224; P. J. Hoffman, The paramorphic representation of clinical judgment, *Psychol. Bull.*, 57, 1960, 116–131; P. J. Hoffman, Cue-consistency and configurality in human judgment, in B. Kleinmuntz (ed.), *Formal Representation of Human Judgment*, 1968, 53–90.

[4] Hoffman, *op. cit.* [in Kleinmuntz], 1968, 61.

[5] P. J. Hoffman, P. Slovic, and L. G. Rorer, An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment, *Psychol. Bull.*, 69, 1968, 338–349; Nancy Wiggins and P. J. Hoffman, Three models of clinical judgment, *J. abnorm. Psychol.*, 73, 1968, 70–77.

[6] N. H. Anderson and Ann Jacobson, Further data on a weighted average model for judgment in a lifted weight task, *Percept. & Psychophys.*, 4, 1968, 81–84.

judgment process, while Hoffman and his co-workers try to determine whether the process is a linear, additive one, or if it is configural.[7] The obvious alternative to this hypothesis is to assume that judgments reflect how *Ss* have learned to utilize the information available to them. This hypothesis demands that we shift our attention from the *subject* side to the *task* side of the problem; for if judgments reflect the ways in which *S* has learned to utilize information, then judgments made with respect to a particular task would reflect the characteristics of that task rather than any general cognitive process within *S*. If this is true, task characteristics have to be as carefully considered as subject characteristics if we are to arrive at an adequate understanding of how humans make judgments. This was precisely the point Brunswik made clear with his lens model (Fig. 1), which pictures *S* as a mirror image of his task.[8]

Testing the hypothesis. The hypothesis that the judgment process is a function of the task characteristics is more conveniently tested in learning experiments than in naturalistic studies of clinicians, for the exact characteristics of the task facing *S* can be made clear in a learning experiment. This is hardly ever possible in nonexperimental studies of clinicians, since even if we know the characteristics of one specific sample of cases for a given task, we cannot reconstruct the general sample of cases that the clinician has encountered in

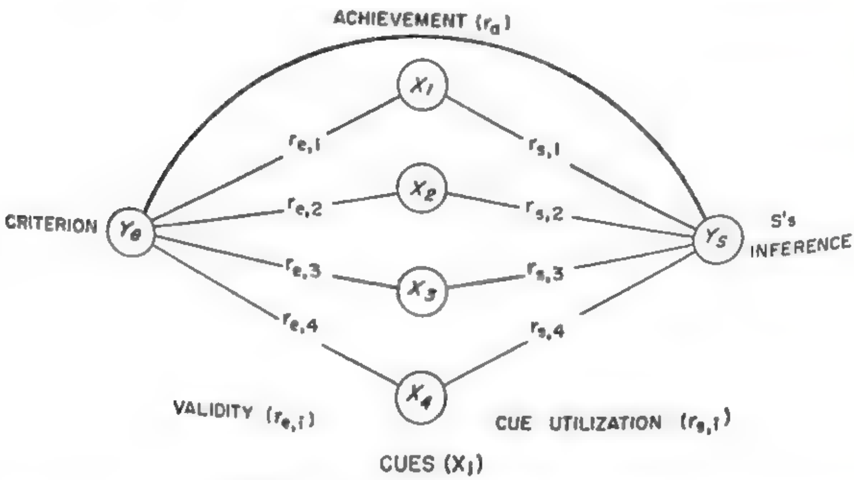


FIG. 1. BRUNSWIK'S LENS MODEL

[7] N. H. Anderson, Averaging versus adding as a stimulus-combination rule in impression formation, *J. exp. Psychol.*, 70, 1965, 394–400; Hoffman, *op. cit.* [in Kleinmuntz], 1968, 53–90.

[8] Egon Brunswik, *Conceptual Framework of Psychology*, 1952.

learning to make his judgments. Utilizing the learning-experiment approach, Hammond and Summers were able to show that humans can learn to utilize nonlinear as well as linear relations for making inferences from a set of cues to a criterion variable.[9] Their results demonstrate that the implication from earlier studies of the process of judgment—the implication that humans are limited to a *linear* use of information—is not generally valid; rather, their results showed that humans can learn to utilize nonlinear relations *if* there are such relations in their task. The second implication from the studies on clinical judgment—the implication that humans are limited to an *additive* combination of information—has remained untested, however. The present study is an attempt to test it by investigating whether *Ss* can learn configural tasks, *i.e.* tasks where a correct judgment cannot be arrived at by an additive combination of cues.

Method

Subjects and design. The *Ss* were 40 undergraduates at the University of Umeå. They were paid to participate in the experiment. The design of the experiment was a 4×8 factorial having four tasks, with different kinds and amounts of configurality, and eight blocks of 50 trials each. The *Ss* were randomly assigned to the four experimental conditions.

Learning tasks. The four tasks required *Ss* to infer the state of a criterion variable from the state of two cue variables. The tasks differed with respect to the relation between the cue values and the criterion values. For Task A, this relation was a linear, additive one; the criterion variable, Y, was a linear function of the sum of the two cue variables, X_1 and X_2, *i.e.* $Y = a + b\ (X_1 + X_2)$. In this task there is, of course, no configural variance. For Task B, the relation was a linear function of the product of the cue variables, *i.e.* $Y = a + b\ (X_1\ X_2)$. This task may either be described as a totally configural one, in that the criterion is a function of the product of the cue values, or as a task with linear as well as configural relations, where the linear relations account for approximately 90% of the variance and the configural relations for the remaining 10%. In Task C, the criterion variable was a linear function of the ratio of the two cues, *i.e.* $Y = a + b\ (X_1 / X_2)$. This task may also be described either as a totally configural one or as a task consisting of both linear and configural components; here, each of the two components accounts for approximately 50% of the variance. In Task D, finally, the criterion values were a linear function of the absolute difference between the cue values, *i.e.* $Y = a + b\ |X_1 - X_2|$. In this task, all of the variance in the cue-criterion relations is configural, in the sense that the correct answer can not be arrived at by a linear, additive use of the cues; the multiple correlation between cues and criterion is $R = .00$. Thus, the four tasks could be ordered with respect to *type* of configurality as well as *amount* of configurality. These two variables

[9] Hammond and Summers, *loc. cit.* (see n. 3 above.)

were confounded, as they must be when configurality is varied. Since the problem investigated in the present experiment was whether *Ss* can learn configural relations, it was more appropriate to sample various types of configurality than to vary the amount of a specific kind of configurality. Using one particular form of configurality would have unduly restricted the results.

For each task, 400 trials were constructed by using all 400 possible combinations of 20 lengths of lines as cue values for both cues. The cue values varied in steps of 1 cm. from 1 through 20 cm. The criterion variable was also a length of a line, varying from 10 to 250 mm. The task was presented to the *Ss* in four booklets. Each booklet contained the cue and criterion values for 100 trials.

Instructions. The instructions were the same in all conditions. The *Ss* were told that there was a relation between the cue values and the criterion values and that their task was to learn this relation so that they could predict the criterion values from the cue values. They were not told the nature of the relation.

Procedure. For each of the 400 trials, each *S* (*a*) observed the cue values, (*b*) made his prediction of the criterion by marking off on a response scale on his answer sheet the length he predicted the criterion line to be, and (*c*) turned the page to observe the correct criterion value. All 400 trials were completed in a single session with a brief pause after each 100 trials. The *Ss* were allowed to work through the material at their own pace. The whole experiment took 2½–3 hr. for each *S*.

Results

The data were analyzed by means of the Tucker version of the lens-model equation which was originally developed by Hursch, Hammond, and Hursch.[10] This equation relates the achievement (r_a) of each *S* to his learning of the linear and nonlinear, or configural, aspects of his task:

$$r_a = GR_eR_s + C\sqrt{1 - R_e^2}\sqrt{1 - R_s^2}\ ,$$

where r_a is the correlation between the predictions made by *S* and the criterion values, G is the correlation between the variance accounted for by a linear multiple-regression equation in the task and that accounted for by a linear multiple-regression equation in *S*'s response system, R_e is the multiple correlation between the cue values and the criterion values, R_s is the multiple correlation between the cue values and the *S*'s responses, and C is the correlation between the variance unaccounted for by a linear multiple-regression equation in the task system and that unaccounted for by such an equation in *S*'s response system. In this equation, G is a measure

[10] L. R. Tucker, A suggested alternative formulation in the development by Hursch, Hammond, and Hursch and by Hammond, Hursch, and Todd, *Psychol. Rev.*, 71, 1964, 528–530; Carolyn Hursch, K. R. Hammond, and J. R. Hursch, Some methodological considerations in multiple-cue probability studies, *Psychol. Rev.*, 71, 1964, 42–60.

of how well an S learns to utilize the linear aspects of his task; C, of how well he learns to utilize the nonlinear or configural relations. For each S and block of 50 trials, r_a, G, C, and R_s^2 were computed.

Achievement. After a transformation to Fisher's z, the achievement scores were analyzed by a 4 (tasks) × 8 (blocks) analysis of variance (ANOVA) with repeated measurements on the second factor. The results of this analysis indicated that there were significant differences between Conditions with respect to the overall level of achievement [$F\ (3/36) = 9.65, p < .01$], as well as a Blocks effect [$F(7/252) = 18.80, p < .01$] and a Condition × Blocks interaction [$F\ (21/252) = 2.40, p < .01$]. The Condition × Blocks interaction was due to the fact that the rate of learning was different in the four conditions (see Fig. 2). While there were large differences with respect to r_a in the first blocks of the experiment, these differences decreased as a function of blocks. The difference between the four groups in the last block was not significant ($p > .05$), indicating that the Ss learned their tasks equally well by the end of the training. Thus, configural relations influenced the *rate* of learning but not the *terminal level* of performance.

Linear aspects. Only in Tasks A, B, and C were there systematic linear relations between cues and criterion. The G scores obtained for these groups were subjected to Fisher's z transformation and analyzed by a 3 (conditions) × 8 (blocks) ANOVA with repeated measurements on the second factor. The results of this analysis showed that there was a reliable difference between Conditions [$F\ (2/27) = 14.63, p < .01$], as well as a Blocks effect [$F\ (7/189) = 7.57, p < .01$] and a Conditions × Blocks interaction [$F\ (14/189) = 2.70, p < .01$]. The loci of these effects are made clear in Fig. 3, which shows G as a function of blocks of trials for the three groups. As Fig. 3 shows, all three groups learned the linear aspects of their tasks, although the learning was much slower for Group C than for the other two groups. Thus, the introduction of configural relations seems to have retarded the learning of the simpler, linear aspects of the task.

Configural aspects. For the three groups—B, C, and D—having systematic configural variance, a 3 × 8 (three conditions and eight blocks of trials) ANOVA with repeated measurements on the second factor was performed on the C scores. The C scores were transformed to Fisher's z before the ANOVA. The results of this analysis

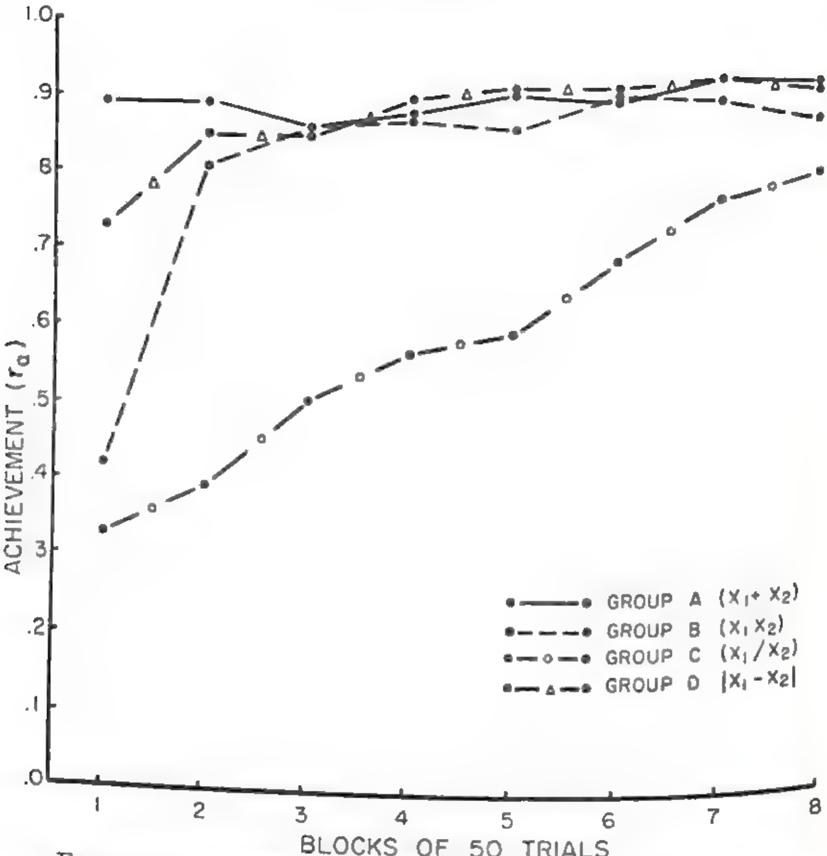


FIG. 2. ACHIEVEMENT (r_a) AS A FUNCTION OF BLOCKS OF TRIALS FOR CONDITIONS A–D

indicated that there was a significant difference between conditions with respect to the overall level of C [F (2/27) = 23.40, $p < .01$], as well as a Blocks effect [F (7/189) = 14.70, $p < .01$] and a barely significant Conditions × Blocks interaction [F (14/189) = 1.80, $p < .05$]. These effects are illustrated in Fig. 4. As can be seen by comparing Figs. 3 and 4, *S*s learned the configural aspects of the task much more slowly than they learned the linear aspects. They also learned the configural aspects less well than they learned the linear aspects, at least for the particular kinds of configurality used in this experiment.

The results from the studies on judgment referred to in the introduction to this paper, as well as the model for functional learning proposed by Björkman,[11] suggest one possible explanation for this

[11] Mats Björkman, Studies in predictive behavior: Explorations into predictive judgments based on functional learning and defined by estimation, categorization, and choice, *Scand. J. Psychol.*, 6, 1965, 129–156.

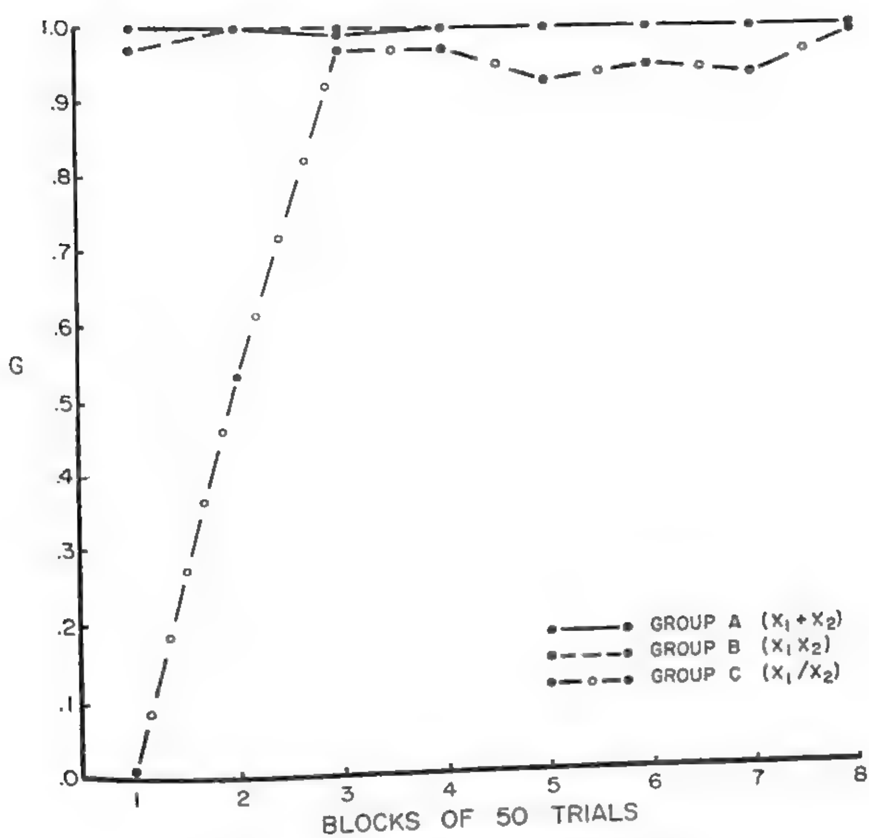


FIG. 3. G AS A FUNCTION OF BLOCKS OF TRIALS FOR CONDITIONS A, B, AND C

result; namely, that the *S*s have a highly linear response system which prevents them from utilizing the configural aspects of the task. To test this hypothesis, the ratio of R_s^2 (the multiple correlation between the cue values and the responses, which is one indicator of the amount of linearity in an *S*'s response system) to R_e^2 (the multiple correlation between cue and criterion values) was computed for each *S* and block for those three groups having systematic linearity in their tasks, *i.e.* for Groups A, B, and C. Fig. 5 shows the mean R_s^2/R_e^2 as a function of blocks for these three groups. As can be seen there, no group had a higher degree of linearity in the response system than there was in the task system. *S*s in Group D (which had a totally configural task) had some linearity in their response systems, although the multiple correlation between cues and responses was generally very low, with a mean across *S*s and blocks of $R_s^2 = .05$. Thus, there is nothing in these data to support

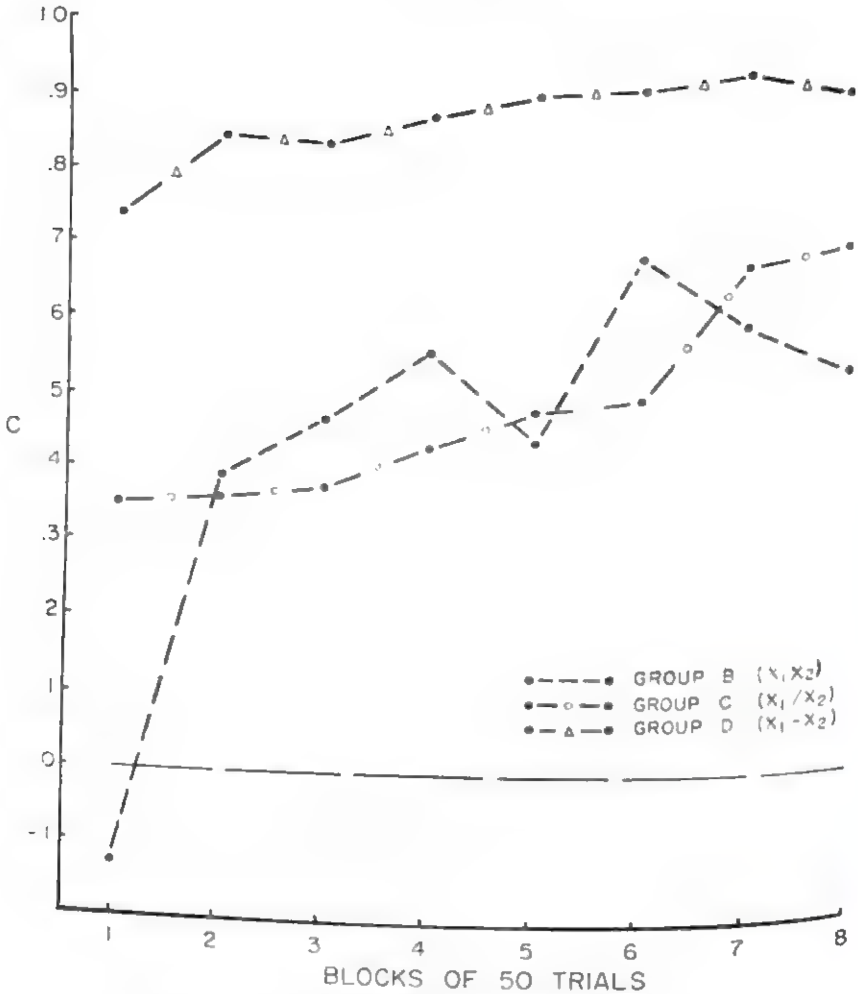


Fig. 4. *C* as a Function of Blocks of Trials for Conditions B, C, and D

the hypothesis that *Ss* tend to have response systems more linear than the task system or that the linearity of their response systems prevents them from utilizing the configural aspects of their tasks.

Discussion

The results of this study show that humans are able to learn to utilize configural cue-criterion relations if there are such relations in the task. The learning of these configural relations was much slower and less efficient than the learning of the linear, additive relations, but the amount of learning was impressive, considering the complexity of the relations used in the learning tasks.

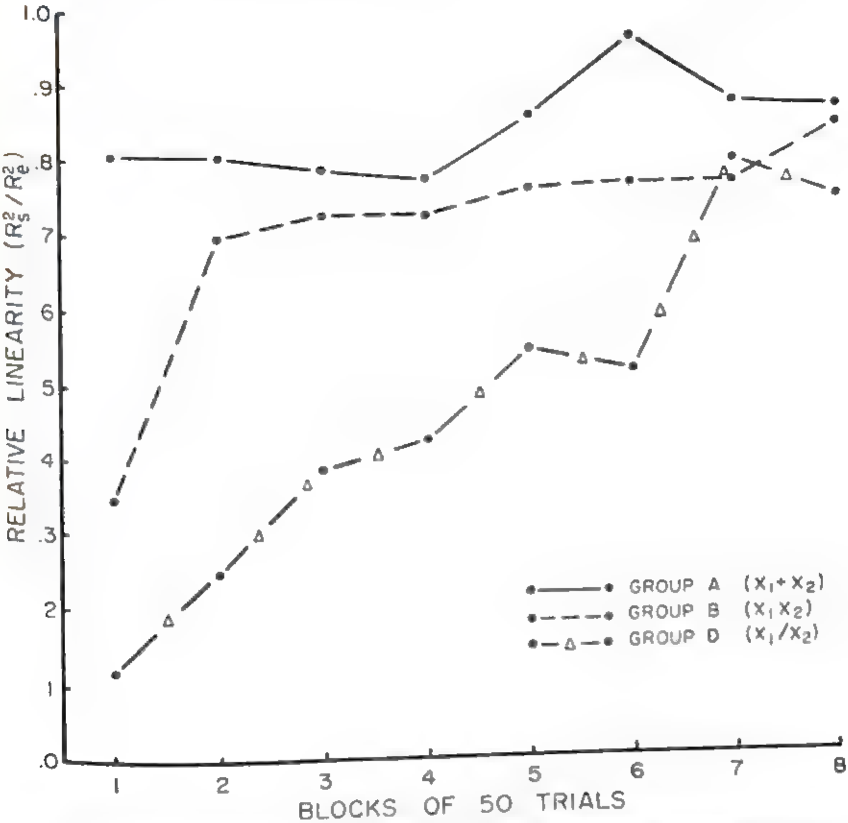


Fig. 5. Relative Linearity (R_s^2 / R_e^2) as a Function of Blocks of Trials for Conditions A, B, and C

There are two obvious limitations to these results. First, they were obtained with a highly simplified judgment task which had only two cues. In the typical inference task (*e.g.* clinical inference), there are considerably more than two sources of information, although, of course, this is not to say that the clinicians use all the sources available to them for their judgments. In fact, there is evidence that they use very few cues.[12] Second, there was no uncertainty in the relations between the cue and criterion values. In a clinical task, the various sources of information typically have a limited validity, which makes the relations between the cues and the criterion uncertain. (Indeed, it may very well be that uncertainty induces linearity.) Within these limitations, however, the results show that humans are by no means *limited* to processing infor-

[12] R. N. Shepard, On subjectively optimum selections among multi-attribute alternatives, in M. W. Shelley and G. L. Bryan (eds.), *Human Judgments and Optimality*, 1964, 257–281.

mation in a linear, additive way: if there are configural relations in the task, the *S*s do, to a considerable extent, learn to utilize these relations for their judgments. Thus, the characteristics of the cognitive process are indeed dependent on the characteristics of the cognitive task.

It is important to note, in this context, that the tasks employed by most investigators in the area of clinical inference have required *S*s to make inferences from test data and that psychological tests are constructed to have linear, additive relations between test scores and criterion scores. Provided the test constructors have been successful, there is very little reason then to expect any configurality in the response systems of the *S*s faced with these tasks. (For demonstration of how little the introduction of configural predictor variables adds to the predictive validity of the MMPI, a favorite task in investigations of the judgment process, see the papers by Wiggins and Hoffman and by Goldberg.)[13] Consequently, it seems reasonable to assume that the results from earlier investigations of the judgmental process have shown the process to be linear and additive because the tasks used have been linear and additive rather than because of any process in *S*s that might have these characteristics.

It must be remembered, however, that a linear multiple-regression equation approximates a number of configural and nonlinear relations.[14] Thus, even if there are configural relations in the task, *e.g.* in the relations between MMPI scores and criterion scores, the analysis of the task by means of multiple regression statistics is likely to show that the task is highly linear and additive. The same holds, of course, for any analysis of the response systems of *S*s by such methods. The same difficulties that face the psychologist trying to analyze the relations between cues and judgments or between test scores and criterion scores by means of multiple regression statistics, also, of course, face a judge trying to find anything but linear, additive relations in his task. If he holds the hypothesis that the relations are linear and additive, his hypothesis has, in most cases,

[13] Wiggins and Hoffman, *loc. cit.* (see n. 5 above); L. R. Goldberg, Diagnosticians versus diagnostic signs: The diagnosis of psychosis versus neurosis from the MMPI, *Psychol. Monogr.*, 79, 1965 (No. 602).

[14] B. F. Green, Jr., Descriptions and explanations: A comment on papers by Hoffman and Edwards, in B. Kleinmuntz (ed.), *Formal Representation of Human Judgment*, 1968, 91–98; Hoffman, *op. cit.*, *Psychol. Bull.*, 57, 1960, 116–131; D. B. Yntema and W. S. Torgerson, Man-computer cooperation in decisions requiring common sense, *IRE Trans. prof. Group hum. Factors*, 1961, HFE 2(1), 20–26.

an extremely low probability of ever being disconfirmed; and he would, in fact, be performing near the optimal level. Therefore, as Carroll pointed out, a linear, additive response system would be adequate for most practical purposes.[15]

As can be seen from the present data, however, humans are not limited to processing information this linear, additive way, although the linear, additive aspects of the tasks are learned more rapidly and to a much higher degree than the configural aspects. Carroll, Hammond and Summers, and Summers and Hammond have previously shown that nonlinear relations are learned less efficiently than linear relations.[16] While these results probably are to some degree a function of the properties of the human information-processing system, it must also be remembered that a statistician, equipped with all the tools of his trade, would require more information (more trials) to determine adequately the nonlinear aspects of a task than the linear aspects, especially in a situation where the criterion was less than perfectly correlated with the cues. Furthermore, it might be, as Björkman has suggested, that linear relations (or good approximations to linear relations) are much more common than nonlinear relations in the *S*s' ecologies, making the difference in rate of learning between linear, additive relations and nonlinear, configural relations a simple case of transfer of training.[17] Thus, the result that linear, additive relations are more easily learned is, probably, a function of two things: the cognitive capacity of the learner (either innate characteristics or transfer effects, or both) and the fact that, all things being equal, nonlinear and nonadditive relations require more information to be adequately detected.

The results on the R_s^2/R_e^2 ratio present another kind of evidence against any hypothesis of a single judgment process. Although the learning of the linear aspects of the tasks was almost perfect, the R_s^2/R_e^2 ratio stayed considerably below unity. Thus, even though *S*'s linear rule was correct on the average, he applied the rule inconsistently. Inconsistency appears to be a general feature of human inference behavior, as a number of studies on multiple-cue probabil-

[15] J. D. Carroll, Functional learning: The learning of continuous functional mappings relating stimulus and response continua, *ETS Res. Bull.*, RB-63-26, 1963.

[16] Carroll, *loc. cit.*; Hammond and Summers, *loc. cit.*; D. A. Summers and K. R. Hammond, Inference behavior in multiple-cue tasks involving both linear and nonlinear relations, *J. exp. Psychol.*, 71, 1966, 751–757.

[17] Mats Björkman, Predictive behavior: Some aspects based on an ecological orientation, *Scand. J. Psychol.*, 7, 1966, 43–57.

ity learning and clinical inference have shown.[18] Some recent data indicate that this inconsistency is due to the fact that *S*s tend to use a number of rules for their predictions interchangeably during the course of an experiment.[19] Thus, we cannot expect to find a process of judgment operating according to some fixed rule. Instead, we must expect to find that the judgments are made according to a number of rules in the same experiment.

Consequently, we cannot conceive of the human judge as a machine, operating according to a fixed rule which is independent of the nature of the input data. Instead, we have to picture him as a highly adaptable system, able to utilize information in many different ways depending on the requirements of the task. He also has, however, the characteristic of inconsistency; *i.e.* he tends to use a number of interchangeable rules for his predictions. It seems reasonable to assume that not only the nature of the rules used by the judge but also the number of different rules and the rate at which he changes his rules depend on the characteristics of the task. If this picture of the human judge is true, it does not seem to be very fruitful to continue to study the nature of the judgment process by analyses of the relations between the input information and the judgments. The characteristics of the judgment task will have to be as carefully considered as the characteristics of the judge. This emphasizes in a striking way Brunswik's basic dictum that in order to understand the nature of the organism, we have to understand the nature of its environment as well.

Implications for clinical prediction. The present data demonstrate that there is in fact some validity to the clinician's claim that he can use information in a configural fashion. On the other hand, there is nothing to prevent the actuary from using configural predictor variables in his regression equations; the statistical method is not inherently limited to a blind application of linear regression statistics to data. There is, however, some doubt that the introduc-

[18] L. W. Dudycha and J. C. Naylor, Characteristics of the human inference process in complex choice behavior situations, *Organ. Behav. hum. Perform.*, 1, 1966, 110–128; K. R. Hammond, Carolyn Hursch, and F. J. Todd, Analyzing the components of clinical inference, *Psychol. Rev.*, 71, 1964, 438–456.

[19] Hiroshi Azuma and L. J. Cronbach, Concept attainment with probabilistic feedback, in K. R. Hammond (ed.), *The Psychology of Egon Brunswik*, 1966, 258–276; B. Kleinmuntz, The processing of clinical information by man and machine, in B. Kleinmuntz (ed.), *Formal Representation of Human Judgment*, 1968, 169.

tion of nonlinear, configural predictor variables would really increase the predictive power of these equations.[20]

Since configural and nonlinear predictor variables can, in fact, be used by the actuary, the main difference between clinical and statistical prediction is not in the kinds of rules used by the two methods but in the consistency with which the rules are being used. While the actuary can use his rules with perfect consistency, the available evidence indicates that the human judge cannot. It can easily be shown that the accuracy of the predictions is a direct function of the consistency with which the prediction rules are applied.[21] Thus, the much criticized inflexibility of the statistical method is an asset rather than a liability. In fact, clinicians would do well to follow Goldberg's suggestion to substitute for themselves regression equations describing their utilization of the input information.[22]

SUMMARY

It was hypothesized that the characteristics of the human cognitive process are dependent on the characteristics of the cognitive task. The general finding, in studies of clinical inference, that humans use information only in a linear, additive way might, therefore, be due to the fact that linear, additive tasks have been used rather than to the characteristics of the human inference process. A learning experiment comparing the learning of linear, additive cue-criterion relations with that of configural relations showed that *Ss* were able to learn configural relations when there were such relations in the task, giving support to the hypothesis. It was concluded that the usual methodology in studies of human judgment, which is to analyze only the relations between the input information and the judgments, is inadequate, and that it is necessary to consider the characteristics of the task as carefully as the characteristics of the *S* to arrive at an understanding of the nature of cognitive processes.

[20] Hoffman, *op. cit.*, 1968, 64.

[21] Hursch, Hammond, and Hursch, *loc. cit.* (see n. 10 above).

[22] L. R. Goldberg, Man vs. model of man: A rationale, plus some evidence for a method of improving on clinical inferences, *Psychol. Bull.*, in press.

CONTINUOUS TACHISTOSCOPIC PATTERNS REPRODUCED WITHOUT LATERALITY EFFECTS

By E. RAE HARCUM, College of William and Mary

This experiment tested the hypothesis that the usual hemifield differences in perceptual accuracy for binary tachistoscopic patterns bisected by fixation are produced by a perceptual scanning of the pattern from end to end, causing one end to be favored by a primacy effect. The patterns used here were called continuous because they extended so far to the left or right of the fixational center that the ends were beyond the range of effective vision. Since there were, in effect, no ends from which an end-to-end visual scanning could begin, the only physical reference point for a sequential perceptual process was the fixation point, from which the hypothesized scanning has no consistent directional tendency. Thus, the design of the experiment was intended to prevent the usual consistent directional scanning of patterns, consequently eliminating laterality effects in the perception. The perceptual results with such patterns should be the same as with presentation of all elements entirely to either the left or right of fixation in separate exposures, *i.e.* should show equality of errors to left and right.[1]

A previous experiment, using the same approach as the present one, failed to eliminate the end-to-end scanning, even with 17-element stimulus patterns covering 7.9° of visual angle.[2] The 17-element patterns revealed generally poor accuracy of reproduction for the elements around fixation, along with most accurate reproduction of the leftmost element within the pattern. Therefore, a presumed perceptual primacy effect favored the left end of the stimulus

* Received for publication April 17, 1969. The investigation reported was supported by a PHS research grant (HD-00207-09) from the Institute of Child Welfare and Human Development, Public Health Service. A preliminary report of this study was included in a previous paper: E. R. Harcum, Lateral dominance as a determinant of temporal order of responding, in M. Kinsbourne (ed.), *Hemispheric Asymmetry of Function*, in press. The author thanks Alan P. Brown for assistance with all phases of the experiment and with the tabulation of data.

[1] D. S. Camp and E. R. Harcum, Visual pattern perception with varied fixation locus and response recording, *Percept. mot. Skills*, 18, 1964, 283–296.

[2] E. R. Harcum, Interactive effects within visual patterns on the discriminability of individual elements, *J. exp. Psychol.*, 68, 1964, 351–356.

pattern rather than the elements about fixation. In the present experiment there were more elements (28) in the pattern, with the end elements falling at positions of greater retinal eccentricity. There were two predictions: one negative and one positive. The negative prediction was that the usual hemiretinal differences in accuracy of reproduction for patterns bisected by fixation would not be obtained with these continuous patterns, regardless of viewing eye. The positive prediction was that the minimum errors would appear for the elements closest to fixation, with symmetrical gradients to the left and right, regardless of viewing eye or with binocular viewing, because the numbers of errors to the left and right of fixation would be highly correlated, positively, within each observation.

Method

Apparatus, stimuli, and procedures were similar to the ones used previously.[3] Binary patterns were typewritten horizontal rows of zeros on white stimulus cards, with various zeros colored with black ink. Each element was 13.2′ through the horizontal diameter, with spaces between elements subtending 17.6′. Total pattern subtended about 13.4°. Field luminance was 2 ftL., and exposure duration was 0.15 sec. There were 28 blackened or open binary elements, arranged to produce 40 different stimulus patterns. Fourteen elements appeared on each side of fixation, with each element position blackened equally often and with seven elements blackened on each side of fixation in each exposure. A fixation cross registered with a point between the two central elements.

Each of 24 paid college-student *O*s was tested for eye dominance and for hand dominance before the experiment began, and then given 10 practice observations. Conditions of left-eye, right-eye, or binocular viewing were counterbalanced for different *O*s. The 40 stimulus patterns within conditions appeared in random order.

O's fixational accuracy was maintained only by instruction; he was instructed to depress a telegraph key, which initiated the exposure, when he was properly fixated on the fixational cross. After each exposure of a pattern, *O* attempted to mark on a prepared data sheet those element positions corresponding to the filled elements in the pattern. The recording sheets contained horizontal rows of 28 typewritten zeros, with a reference marker between the two middle elements. An error was scored each time *O* marked an open element in the stimulus pattern as filled or left a blackened element unfilled.

Results and Discussion

Overall effects. The percentages of errors at each element position are shown in Fig. 1 for three viewing conditions. Since accuracy reached a chance level (50% errors) at about four elements from fixation in both directions, *O*s were obviously not discriminating the

[3] Harcum, *op. cit.*, *J. exp. Psychol.*, 68, 1964, 351–356.

end elements. As predicted, the distributions appeared symmetrical for each viewing condition, with minima near fixation. This gross difference from the usual shape of the distribution of errors—the usual distribution showing minimum errors to the left of fixation and maximum to the right of fixation for most *Os*—supports the present hypothesis in a rather dramatic way.

The distributions in Fig. 1 are similar to those Klemmer obtained using binary lights and poststimulus cues, in that the best performance was for the elements about fixation.[4] Klemmer's results were different, however, in that he found better performance for elements on the extreme left than for those on the extreme right. Since with his 21-element patterns the size of the visual angle subtended by the entire pattern was not a significant variable, Klemmer also concluded that the factors which determined the distributions of errors were perceptual rather than sensory, consistent with the present position.[5]

The results shown in Fig. 1 did not, unfortunately, provide a critical test of the complete implications of the present hypothesis, which stated that the absence of significant hemifield differences is caused by equal performance to the left and right of fixation *within each exposure* for each *O* rather than by differences in opposite directions from exposure to exposure. A finer-grained analysis of these results was needed to test the former possibility.

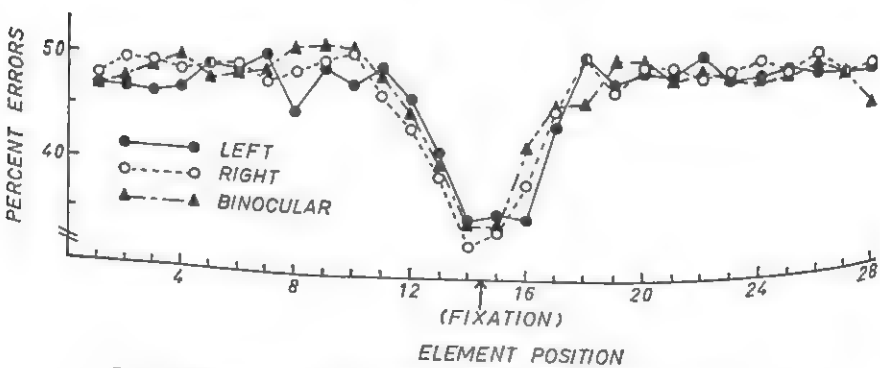


Fig. 1. Percentage of Errors at Each Element Position under each Viewing Condition

[4] E. T. Klemmer, Perception of linear dot patterns, *J. exp. Psychol.*, 65, 1963, 468–473.

[5] E. R. Harcum, Reproduction of linear visual patterns tachistoscopically exposed in various orientations, College of William and Mary, 1964; E. R. Harcum, Parallel functions of serial learning and tachistoscopic pattern perception, *Psychol. Rev.*, 74, 1967, 51–62.

Analysis by single exposures—Selection of data. An analysis was performed on single observations of individual *O*s, but only for data selected by rather stringent requirements. First, only the data from the four element positions immediately on either side of fixation were used. The *O*s showed better than chance performance for these eight elements only, as is indicated by the results plotted in Fig. 1. Second, only those observations in which the stimulus pattern consisted of exactly two filled elements on each side of fixation were used. This meant that the data from 16 patterns—in which three of the elements on one side of fixation were filled, and only one element on the opposite side was filled—were not used. Although such patterns were necessary for the task, they would result in a lateral bias for each exposure. If *O* failed to respond at all, with such patterns he would make three errors on one side and only one error on the opposite side. This bias was counterbalanced over all exposures by the design of the experiment, of course, permitting symmetry such as is seen in Fig. 1. Third, only the data from observations for which *O* overtly marked at least one of the eight elements were used. This ensured that *O* could not produce perfectly equal performance to left and right merely by failing to respond at all; it actually potentially biased the results toward producing unequal hemifield performance, because when *O* marked just one element, it was impossible for him to be equally accurate for elements on right and left. This requirement had the additional advantage of eliminating responses in which *O* obviously was not following instructions. For example, one *O* was probably cheating on fixation in later stages of the experiment. For his last two viewing conditions the minimum of errors fell at the leftmost element, with no obvious dip in errors near the nominal fixation point. This argument is circular, of course, inasmuch as there was no way in this experiment of determining *O*'s integrity of fixation, and this *O* did not admit to cheating. However, a different *O* did, in fact, have to be replaced in the experiment because he became frustrated by this very difficult task, and admitted to cheating on fixation.

By the above criteria there were not sufficient data for two *O*s to permit their use; consequently, the results of these *O*s were not included in the fine-grained analysis. Since the data from these *O*s were also suspect on the basis of the subjective inspection described above, supporting the outcome of the more objective criteria, the loss of these *O*s was probably not detrimental to the interpretation of the results of this experiment.

—*General results of analysis.* Summaries of the analysis by single observations are shown in Tables I, II, and III for left-eye, right-eye, and binocular viewing respectively. These tables represent matrices in which the rows give the number of errors made for the four elements to the left of fixation on that observation and the columns give the number of errors made for the four elements to the right of fixation on that same observation. The cells on the diagonal from upper left to lower right, therefore, represent the cases in which the number of errors to the right and left were equal for the observation. The distance of each cell from this diagonal indicates the magnitude of the difference in errors between left and right of fixation for the individual exposure. The data in each cell are the mean proportions of the observations for each *O* in which a given difference was found. The data of two *O*s were not used in these tables because the numbers of usable observations were too small, as mentioned above. One of these *O*s had no usable observations for one of the viewing conditions, and the other had only two usable observations for a condition. Each of the 22 *O*s whose data appear in the averages in the tables did produce at least 17 usable observations under each condition, however.

The results in Tables I, II, and III suggest two conclusions. The first is that there were no consistent hemifield differences for any viewing condition; this is further discussed below. The second conclusion is that the responses clustered about the central cell, which represent two errors on each side of fixation. This is not surprising, since two errors per side represents chance performance and the task was very difficult, as noted previously. This finding did, however, create a special problem in testing the hypothesis that there would be no difference per observation, since the central cell of the matrix represents one of the five ways that a zero difference between hemifields could be obtained. Such is true even though the data in-

TABLE I

Mean Proportion of Observations per *O* Manifesting Given Numbers of Errors to Left and Right of Fixation with Left-Eye Viewing

Errors to left of fixation	Errors to right of fixation				
	0	1	2	3	4
0	.021	.039	.031	.031	.008
1	.060	.081	.107	.043	.012
2	.047	.081	.092	.065	.015
3	.017	.064	.078	.036	.006
4	.006	.010	.022	.019	.010

TABLE II

MEAN PROPORTION OF OBSERVATIONS PER *O* MANIFESTING GIVEN NUMBERS OF ERRORS TO LEFT AND RIGHT OF FIXATION WITH RIGHT-EYE VIEWING

Errors to left of fixation	Errors to right of fixation: 0	1	2	3	4
0	.011	.061	.046	.025	.002
1	.044	.132	.112	.044	.010
2	.040	.093	.091	.068	.012
3	.013	.052	.054	.042	.008
4	.002	.010	.008	.013	.004

cluded only those cases in which *O* made an overt marking response. The effect implies that there would be a bias in the total data toward finding no difference if *O* tended to be generally accurate in performance. An *O* accurate in perception would rarely produce four errors on one side of fixation and hence seldom produce combinations of zero errors on one side and four on the other—a difference of four errors, of course. On the other hand, such an *O* would often produce fewer errors and, consequently, smaller absolute differences. Thus, a fair comparison would require a control for *O*'s level of accuracy. Unfortunately, this would require further data selection, which might so reduce the available information that only a negative result would be possible. If both the positive and negative predictions of this study could be verified from the remaining data, however, then the method should be justified.

—*Analysis controlled for level of performance.* Since the result of two errors per hemifield represents chance performance, then an accuracy level of one error per side would reflect an intermediate, 'threshold,' level of performance. In such a case, the possibility of differences should be greatest because there is neither a floor (*i.e.* chance errors) nor a ceiling (*i.e.* no errors) on performance. Thus, examination of the three cells in which there is a total of two errors

TABLE III

MEAN PROPORTION OF OBSERVATIONS PER *O* MANIFESTING GIVEN NUMBERS OF ERRORS TO LEFT AND RIGHT OF FIXATION WITH BINOCULAR VIEWING

Errors to left of fixation	Errors to right of fixation: 0	1	2	3	4
0	.028	.063	.029	.023	.004
1	.046	.120	.095	.054	.006
2	.021	.110	.115	.062	.004
3	.019	.043	.056	.055	.004
4	.002	.010	.010	.013	.008

for the exposure should answer this question. These three cells include the case in which there were two errors on the left and none on the right (designated 2–0), the case of one error on each side of fixation (designated 1–1), and the case of no errors on the left and two on the right (designated 0–2). Equal frequencies of cases in the three cells would suggest that there are no consistent biases favoring right or left hemifield. It would also imply that the absence of asymmetry is due to equal frequencies of left-to-right, right-to-left, and nonbiasing scans. The present hypothesis predicted, however, not only equal frequencies of 2-0 and 0-2 combinations but also more frequent 1–1 combinations, *i.e.* nonbiasing scans or no scans at all.

The differences in frequencies of observations in 2–0, 1–1, and 0–2 categories were evaluated by Friedman's two-way analysis of variance.[6] The data for some of the 24 *O*s could not be used because there were no observations in any of the three categories. The differences among the three categories were significant for left-eye viewing ($\chi_r^2 = 9.29$, $N = 23$, $p < .01$), for the right-eye viewing, ($\chi_r^2 = 11.9$, $N = 23$, $p < .01$), and also for binocular viewing ($\chi_r^2 = 11.02$, $N = 20$, $p < .01$). In the case of each viewing condition, the 1–1 category showed a much larger sum of ranked frequencies than either the 2–0 or 0–2 categories, supporting the present hypothesis. This conclusion was further supported by testing pairs of differences between the 1–1 category and the 2–0 and 0–2 categories for each viewing condition, using the Wilcoxon matched-pairs signed-ranks test.[7] For left-eye viewing, the 2–0 and 1–1 categories were significantly different ($T = 33$, $N = 17$, $p < .025$), as were the 0–2 and 1–1 categories ($T = 43.5$, $N = 21$, $p < .01$). For right-eye viewing, the differences were also significant for 2–0 vs. 1–1 ($T = 46$, $N = 21$, $p < .01$) and for 0–2 vs. 1–1 ($T = 33$, $N = 21$, $p < .005$). With binocular viewing also, differences were significant for 2–0 vs. 1–1 ($T = 19$, $N = 18$, $p < .005$) and for 0–2 vs. 1–1 ($T = 21$, $N = 18$, $p < .005$). Therefore, there were significantly more cases of equal performance to the right and left than there were cases of biases in either direction.

The question of possible hemifield differences can be settled by comparing the frequencies of 2–0 and 0–2 responses for each viewing condition. The differences were not significant for the left eye ($T = 50$, $N = 16$, $p > .05$), or for the right eye ($T = 56$, $N =$

[6] S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*, 1956.
[7] Siegel, *op. cit.*

15, $p > .05$), or for binocular viewing ($T = 26$, $N = 11$, $p > .05$). Therefore, the evidence indicates that there was no consistent directional bias favoring right or left visual hemifield, even for those observations that did in fact show a difference between hemifields.

Order of report. The conclusion that the perceptual primacy effect does not favor the elements at one end of the pattern but rather those elements nearest fixation was supported by *E*'s observations of *O*s' order of marking the elements. The *O*s typically marked only the elements close to fixation, and they did so in no consistent sequence. This result is contrary to the usual tendency of *O*s to mark elements from the left end toward the right when the patterns have perceivable ends.[8] The fact that overt responses of the *O*s clustered about fixation undoubtedly contributed to the measured perceptual accuracy for those elements. However, Harcum, Hartman, and Smith have shown that the marking sequence is correlated with the scanning sequence when *O* is permitted his option of marking sequences.

Possible effects of dominance. Since there were no hemifield differences in performance in this experiment, there could be no differential effect of cerebral hemispheric dominance or of eye dominance on relative hemifield performance for individual *O*s. Thus, visual-sensitivity differences which might be related to such factors did not exhibit an effect under the conditions of the present experiment. The previously obtained effects of viewing eye for such binary tachistopic patterns exposed across fixation would appear related to mechanisms which influence scanning direction, as Harcum and Dyer concluded, rather than to mechanisms of differential visual sensitivity.[9] This conclusion does not deny the possibility of alternative mechanisms which can produce laterality effects and which can thus be listed under the rubric of dominance.[10] Individual differ-

[8] E. R. Harcum, R. R. Hartman, and N. F. Smith, Pre- vs. post-knowledge of required reproduction sequences for tachistoscopic patterns, *Canad. J. Psychol.*, 17, 1963, 264–273.

[9] E. R. Harcum and D. W. Dyer, Monocular and binocular reproduction of binary stimuli appearing right and left of fixation, this JOURNAL, 75, 1962, 56–65.

[10] M. P. Bryden, Tachistoscopic recognition and cerebral dominance, *Percept. mot. Skills*, 19, 1964, 686; M. P. Bryden, Tachistoscopic recognition, handedness, and cerebral dominance, *Neuropsychologia*, 3, 1965, 1–8; D. Kimura, Cerebal dominance and the perception of verbal stimuli, *Canad. J. Psychol.*, 15, 1961, 166–171; W. A. Winnick and R. L. Dornbush, Pre- and post-exposure processes in tachistoscopic identification, *Percept. mot. Skills*, 20, 1965, 107–113.

ences can obviously be important.[11] The primary problem here is to discover the nature of the actual relevant mechanisms, of which there may be two.[12]

Conclusions. The evidence indicates that the symmetrical distributions in Fig. 1 probably were not the result of counterbalancing unequal errors on different exposures. Instead of fewer cases of equal performance to the right and left, there were significantly more cases of such equal performance than of biases in either direction. Thus, the positive result predicted for this study adds weight to the negative result, which was also obtained. It indicates that the negative result was not obtained merely because the experimental test was not sufficiently sensitive to reveal significant differences. Although a negative cannot be proven, the present data indicate that the test was sufficiently sensitive to have detected a hemifield difference if there had been one.

These symmetrical distributions of errors obtained with patterns bisected by fixation are similar to those usually found for patterns exposed entirely on one side of fixation.[13] This further supports the argument that scanning proceeds from fixation toward the edge of the visual field. The results thus support the hypothesis that a primacy effect in a spatial-temporal perceptual process accounts for the usual lateral differences for multiple-element patterns exposed across fixation.

Summary

The hypothesis was that hemifield differences in perceptual accuracy for tachistoscopic patterns exposed across fixation are produced by a primacy effect in a perceptual process which scans the stimulus pattern from one end to the other. Twenty-four *Os* reproduced horizontal tachistoscopic patterns which extended so far into peripheral vision that the patterns had no perceivable ends, thereby preventing end-to-end scanning. Thus, elements in neither hemifield should have been favored by a perceptual primacy effect. Elements nearest fixation were most accurately reproduced, and no lateral differences were found, either for binocular viewing or for monocular viewing with either eye, supporting the predictions.

[11] Harcum, *op. cit., J. exp. Psychol.*, 68, 1964, 351–356; Harcum, *op. cit.* [in Kinsbourne], in press.

[12] E. R. Harcum, Two possible mechanisms of differential set in tachistoscopic perception of multiple targets, *Percept. mot. Skills*, 25, 1967, 289–304; W. Heron, Perception as a function of retinal locus and attention, this Journal, 70, 1957, 38–48.

[13] D. S. Camp and E. R. Harcum, Visual pattern perception with varied fixation locus and response recording, *Percept. mot. Skills*, 18, 1964, 283–296.

FAMILIAR PATTERNS ARE NO EASIER TO SEE THAN NOVEL ONES

By JOHN S. ROBINSON, Brain-Behavior Research Center, Sonoma State Hospital, Eldridge, California

It is commonly thought that the more familiar a pattern, the easier it is to see; and proponents of such an idea seek to support it by appealing to the 'word-frequency effect,' the inverse relation between the frequency of occurrence of a word in print and the viewing time required for its identification.[1] Studies have appeared recently which oppose this view, attributing threshold differences for frequent and rare configurations to nonperceptual factors (differences in the retention or response-selection requirements for reporting what is seen).[2] They show that, indeed, when control is provided for such factors, differences in perceptibility cannot be demonstrated.[3] Similarly, the data of the present study can be most readily understood if it is assumed (1) that repeated viewing of patterns does not increase their perceptibility and (2) that instances in

* Received for publication May 1, 1969. The research reported was supported by California State Department of Mental Hygiene Grant 67-16-25 and by Grant MH-08832 from the National Institutes of Mental Health. The author would like to thank Dr. Marilyn Milligan for her expert assistance and the Sonoma State College Psychology Department for providing research space.

[1] Davis Howes, An approach to the quantitative analysis of word blindness, in J. Money (ed.), *Reading Disability,* 1962, 140; R. N. Haber, Repetition as a determinant of perceptual recognition processes, in W. Walthen-Dunn (ed.), *Models for the Perception of Speech and Visual Form,* 1967, 207.

[2] J. S. Robinson, L. T. Brown, and W. N. Hayes, Test of effects of past experience on perception, *Percept. mot. Skills,* 18, 1964, 953–956; W. E. Foote and L. L. Havens, Differential effects of stimulus frequency and graphic configuration in free- and forced-choice experiments, *J. exp. Psychol.,* 73, 1967, 340–346; Bruce Earhard, Perception and retention of familiar and unfamiliar material, *J. exp. Psychol.,* 76, 1968, 584–595; Arthur Wingfield, Effects of frequency on identification and naming of objects, this JOURNAL, 81, 1968, 226–234; J. S. Robinson, Paired comparison and sample-matching thresholds for familiar and unfamiliar verbal materials and line patterns (unpublished).

[3] See, however, image-fragmentation studies supporting the view of an effect of familiarity on pattern perception: R. M. Pritchard, W. Heron, and D. O. Hebb, Visual perception approached by the method of stabilized images, *Canad. J. Psychol.,* 14, 1960, 67–77; J. P. McKinney, Verbal meaning and perceptual stability, *Canad. J. Psychol.,* 20, 1966, 237–242. Important methodological and theoretical innovations in this area of research are presented in C. R. Evans, Further studies of pattern perception and a stabilized retinal image: The use of prolonged after-images to achieve perfect stabilization, *Brit. J. Psychol.,* 58, 1967, 315–327.

which there is an apparent increase in perceptibility, as in the word-frequency effect, actually involve a reduction in perception of the stimulus.

'Perceptibility' is sometimes used to refer to the clarity and completeness with which an object is seen, or it may be used to indicate 'identifiability.' Both meanings are illustrated in a tachistoscopic experiment by Haber and Hershenson in which the effects of massed repetition on the perceptibility of very briefly exposed words were studied.[4] A word was flashed repeatedly at intervals of, say, 8 sec.; flash duration was constant but so short that nothing could be seen with only one flash. With repeated trials, however, *Ss* were able first to see fragments of letters, then some complete letters, and finally the entire word with perfect clarity.[5] Words were perceptible in the sense of 'identifiable' in the early stages, when *Ss* could see very little of the stimulus; at a given stage of percept organization, while the percept was still quite incomplete, the more familiar the word, the more perceptible, *i.e.* identifiable, it was—the word-frequency effect. Only in the final stage, after repeated massed trials, did words become perceptible in the sense that they could be seen clearly and completely.

Repetition works in opposite directions in later and earlier stages. In the later stages, with more and more trials, the percept becomes clearer and more complete; in the earlier stages, with more and more 'trials' (*i.e.* greater frequency of occurrence in print), the percept on which word identification is based becomes less and less clear and complete. As familiarity with a word increases, "preattentive"[6] recognition develops: a partial percept—a fleeting, incomplete image of the word—comes to stand for the whole and to provide the basis for its identification. The reality of the perceptual reduction involved in the ready identification of familiar words is attested to by the 'proofreader's error.' Such errors involve overlooking those typist's or printer's mistakes which do not change the aspects of the word (the general configuration, distinctive features) on which preattentive recognition is based.

An attempt was made in the present study to find out whether repeated viewing could, in fact, increase the clearness and complete-

[4] R. N. Haber and M. Hershenson, Effects of repeated brief exposures on the growth of a percept, *J. exp. Psychol.*, 69, 1965, 40–46.

[5] The final stage effect of the massed repetition employed is a short-term one, but it illustrates the kind of qualitative change that might conceivably occur as a long-term consequence of other conditions of repetition.

[6] Ulric Neisser, *Cognitive Psychology*, 1967, 89.

ness with which patterns are seen; or whether any improvement in visual processing time it produced could be explained in terms of preattentive recognition of stimuli, with its accompanying reduction in the amount of the stimulus perceived. The *S*s were given tachistoscopic presentations of initially unfamiliar patterns on a number of successive days. Some of the patterns were seen only once, others repeatedly. The patterns were briefly exposed identical or closely similar pairs of Chinese characters (Figs. 1 and 2) which *S*s had to judge as same or different. The patterns were not as readily encod-

FIG. 1. SAMPLE SAME AND DIFFERENT STIMULUS PAIRS

able as English words or letters, so that *Ss* had to compare the actual visual configurations or their "iconic storage"[7] counterparts, in making their judgments. The stimuli were made large enough so they could be seen easily, but small enough so that multiple fixations were unnecessary; the grounds for this being that when more than one fixation is required, a retention factor is introduced, since earlier features must be remembered while later ones are fixated (and changes of fixation may themselves introduce response artifacts). There was a perfect equating of response requirements for novel (presented only once) and familiar (presented repeatedly) patterns, since no other response than "Same" or "Different" had to be made to either.

The perceptual requirements of Same and Different judgments are not symmetrical: a partial view of a Different pair may be enough for *S* to see a difference, but a complete view of a Same pair is necessary if he is to see the components as the same. If there is improvement in the judging of stimuli that have been seen repeatedly, this asymmetry of perceptual requirements can be used to find out how the improvement occurs. It may be that visual processing of Different pairs is speeded up by the development of selective attending to the differential portions of the configurations, guided by preattentive recognition of repeatedly seen pairs. The process would be similar to the one assumed to underlie the word-frequency effect, with selective attention paralleling word identification. Judgment of Same pairs requires seeing all of both components, so that increased speed of processing could not occur on the same kind of selective-attending basis.

Accuracy and confidence of judgment of Different, but not of Same, pairs, therefore, should increase if preattentive recognition and selective attending are the primary factors involved in changes that take place with repetition. (The increases should be especially evident for those exposure durations which are long enough to allow a single focusing of attention but too short to permit attention shifts.) On the other hand, judgments of both Same and Different pairs should improve if there is an increase in perceptual clarity with repetition. The former change would represent a decrease, the latter an increase, in what is actually seen of the patterns.

[7] Neisser, *op. cit.*, 1967, 15–45.

Method

Subjects. Forty-eight undergraduate *Ss* were divided into six groups. Each group received one of six 24-pair sets of patterns as 'Repeated' stimuli (presented once a day on all five test days) and the other five as 'Unique' (presented only once). A balanced design provided control for the effects of using a particular set as the Unique patterns on a given one of the five test days or as Repeated stimuli.

Stimuli. Same and Different configurations were made out of basic patterns selected from Weiger's *Chinese Characters*.[8] The nine different loci represented by the cells in a 3 × 3 grid were used equally often in modifying the characters (Fig. 2). Basic and variant Same pairs, each accompanied by one of the related Different pairs, were put in separate sets. There were 12 Same and 12 Different pairs in each of the six sets of patterns.

Apparatus and procedure. The pattern pairs were viewed at a distance of 126 cm. in a Scientific Prototype Model GB Tachistoscope; they subtended 1.8° of visual angle horizontally and .9° vertically. One channel was used for pre- and postexposure fields, another to present the patterns. Brightness in both was 14 ftL. An appropriate range of exposure durations—5, 25, 100, 500, 1000 and 2000 msec.—was established in a preliminary series. Repeated/Unique, Same/Different, and exposure-duration differences were balanced over the *Ss*' 48 daily trials, and presentations varied randomly from trial to trial with respect to all three. The *Ss* were told how to view the stimuli in the tachistoscope and were given sample Same and Different configurations to examine; also, they were told how the stimuli and exposure durations would vary and that the variations would appear in random sequence. They were in-

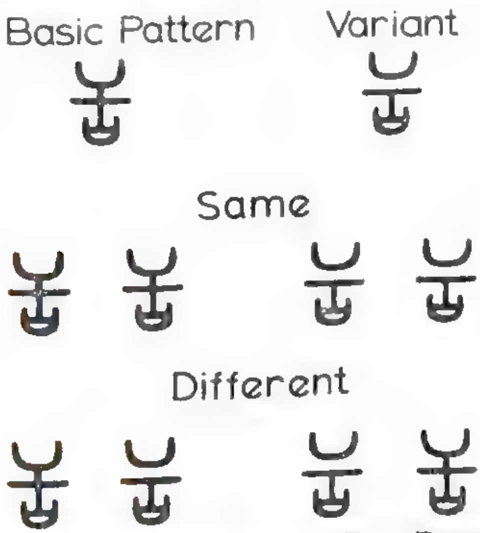


Fig. 2. Same and Different Stimulus Pairs Derived from a Basic Pattern and Its Variant

[8] L. Weiger, *Chinese Characters*, 1965, 397–566. Patterns were drawn in india ink on bond paper with a Rapidograph No. 4 pen and reproduced by Diazo print method. Copies available from the author.

structed to judge the patterns as same or different and to report how confident they were of their judgment by using a scale of 1 to 3, with 1 indicating "doubt," 3 "certainty." They were not told if their judgments were correct. A trial lasted about 25 sec.

Results

Responses to Different-Repeated patterns were the only ones to show a significant, or even consistent, change (see Fig. 3). Comparison of *Ss'* Day 1 scores with their mean scores for Days 2–5[9] showed that both accuracy and confidence increased significantly: $p < .001$ for both. Comparison of mean-change scores (mean score, Days 2–5, minus Day 1 score) for Different-Repeated and Different-Unique patterns also yielded significant values: accuracy, $p < .005$; confidence, $p < .001$.

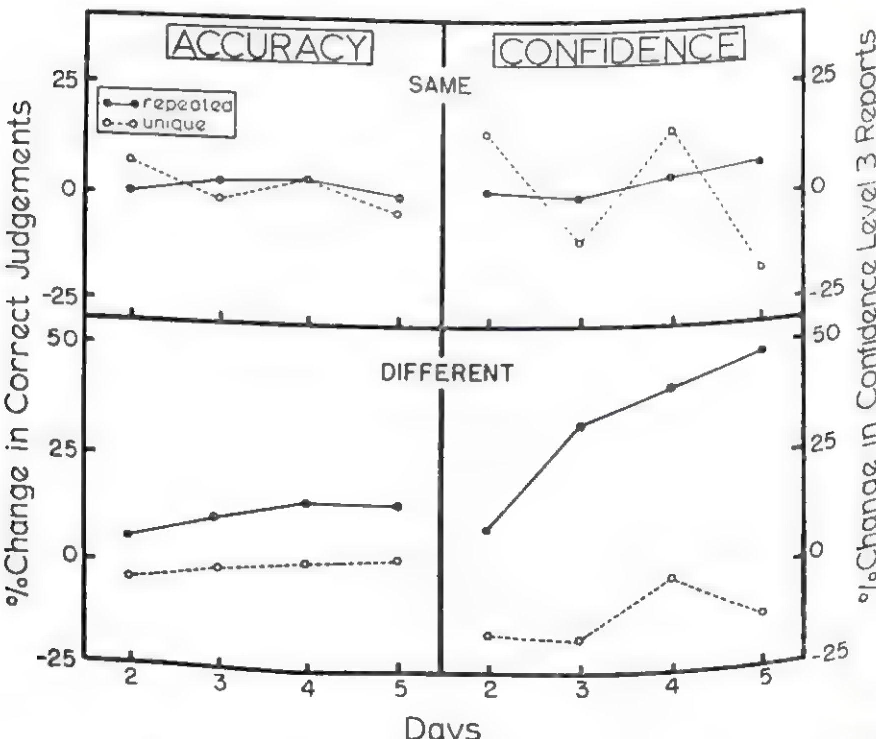


Fig. 3. Changes in Number of Correct Judgments and Confidence Level 3 Reports on Days 2–5 Expressed as Percentage of Day 1 Values (Points based on a total of 576 trials [48 Ss × 12 Trials])

[9] Wilcoxon's paired-replicates method was used in tests involving comparison of pairs of related scores; Friedman's analysis of variance was used to test the significance of overall differences among the scores for the six separate exposure durations. One-tailed probabilities are given.

The six exposure durations did not contribute equally to the improvement shown in response to Different-Repeated patterns: differences among their mean-change scores were highly significant, $p < .001$ for both accuracy and confidence. Improvement was significant at intermediate durations only (see Fig. 4). For comparison of Repeated and Unique scores at the 25-, 100-, and 500-msec. durations, $p < .025$ .001, and .005 respectively for accuracy scores, and .001, .001, and .04 for confidence scores.

Discussion

There is no evidence that familiarity makes patterns easier to see, since there was no improvement in the accuracy of judgment of Same pairs. The fact that judgments of Different pairs did show significant improvement while judgments of Same pairs did not may be a reflection of the differential perceptual requirements for Same- and Different-pair judgments discussed above. That is, a partial view of a Different pair can mediate correct judgment, but both components must be seen in their entirety for correct Same judgments. Repeated presentation of the Different pairs enables *S* to find which part to attend to and which to neglect (with consequent speeding up of visual processing time). As outlined above, preattentive recognition serves as a guide for selective attending to the differential aspects of the different pairs. Further support for this interpretation is provided by the finding that improvement was significant at only the intermediate durations. That is, with a brief glimpse of previously seen patterns, *S* can speedily 'zero in' on the critical differential part of the configuration. Nothing is seen if the exposure is too short, and the longer exposures deprive familiar patterns of any advantage over novel ones, since *S* has time to examine all parts of either type of configuration.

The basic conclusion to be drawn here is the same as the one reached in the studies (cited above) which reported that familiarity does not make patterns easier to see. The improvement in Different-pair judgments, however, shows how the mistaken notion of increased perceptibility, as expressed in studies of the word-frequency effect, could arise. With increases in familiarity, ever briefer glimpses of the stimulus mediate preattentive recognition, which cues off selective attending (Different pairs) or identification (words) with ever-greater facility.

The overwhelming importance for visual processing of a readily accessible method of stimulus encodement is illustrated by the large

difference in exposure time required for *seeing* the not-readily-encoded patterns in this study and for *identifying* the stimuli used in studies which employed materials which could be readily encoded (letters, words, pictures of nameable objects). Long exposure durations were required for the *S*s in this study to see the patterns with any degree of accuracy and confidence (Fig. 4), whereas even pictures of objects with infrequently occurring names (*e.g.* gyroscope) in the Wingfield study[10] could be identified in less than 30 msec. With encodable items, guessing can supply the missing parts when the percept is fragmentary. Short duration thresholds do not mean that familiar items can be seen more readily; they simply attest to the fact that percept organization, the actual buildup of contours, does not have to proceed very far at all for identification to occur when some method of encodement is available.

This difference between the relatively long time needed for seeing the unencoded and the short time sufficing for identifying the encoded probably plays an important role in Earhard's Experiment III.[11] His *S*s judged a tachistoscopically presented letter pattern as same or different compared with a second pattern they could examine freely. Earhard found that the more detail there was to examine (*i.e.* the more letters nonmatching standard and comparison terms had in common), the more effect familiarity had on accuracy of judgment. He explained the effect in terms of the difference in amount of detail of familiar (high-frequency words) and unfamiliar (nonsense-letter sequences) items retained from exposures at a given duration. Retention certainly must have played a role, but it is important to note that familiar items could be identified at short exposures and that a fragmentary percept could be supplemented by the inferred remainder of the stimulus when comparison was made subsequently; if a nonsense-letter sequence was presented, however, *S* had to use what little he had actually seen in making his comparison. With nine-letter items being shown for as little as 10–20 msec., such an explanation of the obtained familiarity effect seems especially relevant.

Hochberg's view of perceptual learning receives considerable support from these data. He assumed that the basic perceptual apparatus is inherited and essentially unmodifiable, and that what is called 'perceptual learning' actually involves changes in how attention is

[10] Wingfield, *op. cit.*, 227–228.
[11] Earhard, *op. cit.*, 591–595.

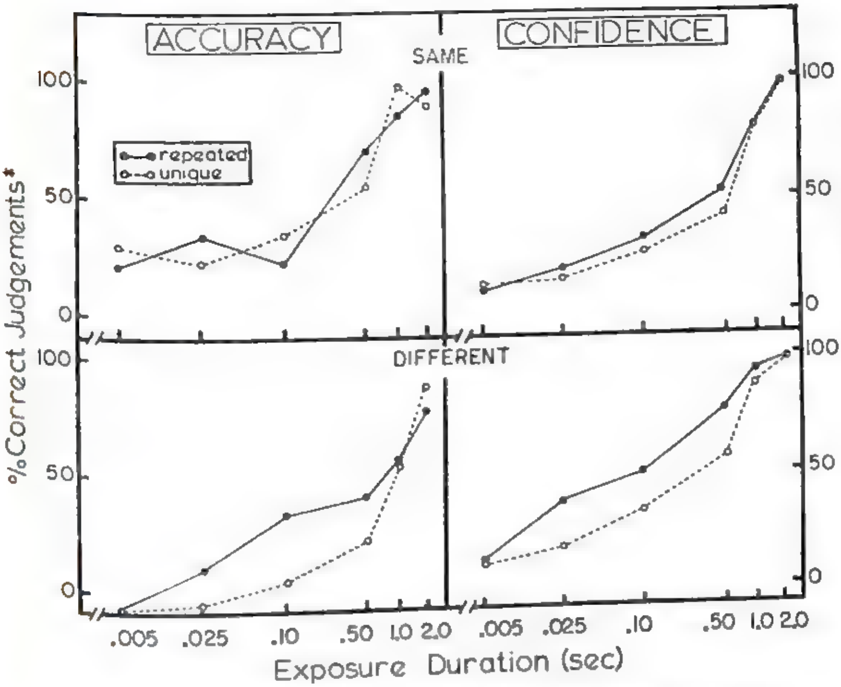


FIG. 4. PERFORMANCE AS A FUNCTION OF EXPOSURE DURATION: REPEATEDLY SEEN PATTERNS AND PATTERNS SEEN ONLY ONCE ('Unique' points based on 6 days [Days 1–5, Unique; Day 1, Repeated] × 48 *Ss* × 2 trials per day = 576 trials; 'Repeated,' on 4 × 48 × 2 = 384 trials; *Percentage values corrected for guessing [% correct − % incorrect]; **Based on trials on which patterns were judged correctly)

deployed and in how what is seen is remembered.[12] On the other hand, the results are much less readily interpreted in terms of Hebb's ideas about perceptual learning, which assume the building up and modifying of basic form-reception mechanisms by experience.[13] The results favoring the Hebbian view of perceptual development[14] may do so because the studies involved maintenance of already organized percepts, whereas in this experiment the focus of interest was on the initial building-up of the contours of the percept. Experience might be expected to play quite different roles in the two processes.

Although repeated exposure to the kinds of patterns used in this study does not make them easier to see, it apparently can make them easier to differentiate. It does so, however, through a process of selective attention to specific details found to be critical, with neg-

[12] Julian Hochberg, In the mind's eye, in R. N. Haber (ed.), *Contemporary Theory and Research in Visual Perception*, 1968, 325–330.
[13] D. O. Hebb, *The Organization of Behavior*, 1949, 60–106.
[14] Pritchard, Heron, and Hebb, *loc. cit.*; McKinney, *loc. cit.*

lect of other nondifferentiating features. Thus, an increase in accuracy of differentiation is accompanied by an overall decrease in what is seen of the patterns.[15]

Summary

Comparison was made of the clearness with which novel and familiar patterns are seen. Young adult *Ss* judged tachistoscopically presented identical or closely similar pattern pairs "Same" or "Different." There was no increase in accuracy or confidence of judgment over a five-day test period for repeatedly presented Same patterns, but there was an increase for Different patterns. Familiarity apparently does not increase the clarity with which patterns are seen, and may actually reduce it: the improvement in judgment of Different pairs, considered in conjunction with the lack of improvement for Same pairs, can best be understood in terms of selective attention to differential aspects of the stimuli—and consequent neglect of their nondifferential features—which develops as the patterns come to be recognized.

[15] Further tests of the interpretation given here could be made by (*a*) obtaining phenomenal reports of relative clearness of perception of nondifferential features of the stimuli as differentiation improved and (*b*) reducing or eliminating improvement by using Different pairs which made selective attention difficult or impossible.

THE ORGANIZATION OF HYPOTHESIS-TESTING BEHAVIOR IN CONCEPT-IDENTIFICATION TASKS

By O. Ernest Andrews, Charles F. Levinthal,
and Harold D. Fishbein,
University of Cincinnati

The major purposes of this paper were twofold: to test extensively the generality and validity of the conclusion by Fishbein, Benton, Osborne, and Wise that in concept-identification tasks, the stimulus dimensions play an important organizational role in the way *Ss* test hypotheses,[1] and to determine the extent to which the number of dimensions and values within each dimension affects selective memory loss from Trial n to Trial $n+1$. In the Fishbein *et al.* study, *Ss* were run in either three-dimension, two-value, or six-dimension, two-value, successive discrimination-learning tasks. A modification of the blank-trials technique of Levine[2] was so employed that on each outcome trial, *Ss* received information about only one hypothesis, *i.e.* one value of one dimension. These authors found that following an error on Trial 1, the probability of *Ss* testing on Trial 2 a hypothesis that lay along the same stimulus dimension as their Trial 1 hypothesis was significantly greater than chance. Does this behavior occur when simultaneous discrimination tasks are used? More importantly, does the dimension continue to organize *Ss'* behavior beyond the first two outcome trials?

The second major set of questions arose from the fact that when the blank-trials technique is utilized, *S* is informed on each outcome trial about both the correctness of the hypothesis he held during the no-outcome trials and the correctness of other hypotheses, *i.e.* the other values of the chosen stimulus. Levine has shown that in

* Received for publication July 5, 1968. O. E. Andrews is now at the State University of New York at Stonybrook; C. F. Levinthal, at the University of Michigan. The authors thank Mrs. O. E. Andrews for assistance in the data analysis. They ask that reprint requests be directed to Dr. Fishbein.

[1] H. D. Fishbein, J. Benton, M. Osborne, and H. Wise, Effects of a "wrong" on the rejection of hypotheses and dimensions in concept learning tasks, *Proc. 74th Annu. Conv. Amer. Psychol. Ass.*, 1967, 45–46.

[2] M. Levine, Hypothesis behavior by humans during discrimination learning, *J. exp. Psychol.*, 71, 1966, 331–338.

four-dimension, two-value tasks, Ss have virtually complete memory from Trial n to Trial $n+1$ for the hypothesis held during blank trials, but from his Fig. 4 it appears that Ss suffer some memory loss for the other hypotheses.[3] Is this a valid observation, and to what extent do variations in number of dimensions and values within a dimension affect this differential memory loss?

Method

Procedure. The Ss were 267 undergraduates randomly assigned to six groups (approximately 45 Ss per group). The three-, four-, or six-dimension, two- or four-value stimuli were constructed on 3 × 5 in. cards. The stimulus dimensions were: position of the geometrical figure, shape of the larger (outside) geometrical figure, color of the larger (outside) geometrical figure (three-dimension cards), number of bars under the figures (four-dimension stimuli), shape of the smaller (inside) geometrical figure, and color of the smaller (inside) geometrical figure (six-dimension stimuli).

The Ss received four training tasks and three test tasks of simultaneous discrimination problems; *e.g.* in tasks using four-value dimensions, four stimuli were shown each trial, and S picked what he thought was the correct one. The Ss were instructed that the correct concept involved one of the values of one of the dimensions, with no other solutions possible. For the third and fourth training tasks, no-outcome (blank) trials were introduced. If S consistently chose a single value (*e.g.* red) on one dimension (*e.g.* outside figures), his pattern of responses on four consecutive blank trials uniquely identified that value (hypothesis). The fourth practice problem and the three test problems consisted of four blank trials alternated with single outcome trials. For the two-value tasks, 24 trials were run, of which Trials 5, 10, 15, and 20 were outcome trials and the remainder blanks. For the four-value tasks, 34 trials were run, of which Trials 5, 10, 15, 20, 25, and 30 were outcome trials. Hence, the hypothesis held by S on his first outcome trial could be identified. All Ss were told "wrong" on the first outcome trial of the test tasks (on each outcome trial, E said only "wrong" or "right"). Following this, E, according to a prearranged random schedule, made one of the remaining nondisconfirmed values the correct one. After the last trial of each task, Ss were informed of the correct solution. Each task was S-paced.

Response measures. A hypothesis (H) was defined as a pattern of responses—on a series of four blank trials and the following outcome trial—which corresponded to exactly one value on one dimension. (The uncapitalized word 'hypothesis' is used, in the following, as a synonym for 'single value of one dimension.' Its capitalization or abbreviation to H implies the pattern of responses on blank trials.) The primary hypothesis (PH) was the H an S used on Block 1 (Trials 1–5). The secondary hypotheses (SHs) were defined as those values of the stimulus, other than the PH, which S chose on Trial 5. For example, on a three-dimension, two-value task, the S may have held the PH of red. On Trial 5, he chose a red square on the left and was told that he was wrong. In this case, left-hand position and squareness were SHs. A primary di-

[3] Levine, *loc. cit.*

mension (PD) was defined as the dimension along which the PH lay; *e.g.* the PD for a PH of left position would be position. The secondary dimensions (SDs) were defined as the dimensions along which the SHs lay. In an *n*-dimensional task there are $n-1$ SDs. Note that PHs, SHs, and SDs are defined relative to performance only during Block 1.

Results

In all analyses, following the example of Levine, the data were pooled across *S*s and tasks for each group. If *S* failed to maintain a consistent H (Hypothesis) on a given block of five trials (the blank trials and the following outcome trial), those data were eliminated from the analyses. Empirically, over all tasks and trial blocks *S*s maintained a consistent H approximately 90% of the time. This compares favorably with Levine's data.[4]

Tables I, II, III, and IV present information bearing on the question of the organization of hypothesis-testing behavior. Table I indicates the predicted and obtained proportions of testing an H on Block 2 (Trials 6–10) that lay along the Block 1 (Trials 1–5) Primary Dimension (PD). The predicted proportions assumed that *S*s either have complete memory of the information received on Trial 5 and do not test disconfirmed Hs or SHs, or have complete absence of memory of Trial 5[5] and randomly select H from the set of tenable hypotheses. As can be seen, under all conditions *S*s tested along the PD to a greater extent than expected by chance. The differences between the obtained and predicted proportions were statistically significant from chance in all cases.

Tables II, III, and IV present data from *S*s run in four-value tasks. Table II shows the predicted and obtained proportions of *S*s testing along their PD on Blocks 1, 2, and 3 (Trials 11–15), and 4

TABLE I

Predicted and Obtained Proportions for Testing along Same Dimension from Block 1 to Block 2

	Two-value tasks		Four-value tasks	
	Predicted	Obtained	Predicted	Obtained
Three-dimension	.33	.43	.33	.48
Four-dimension	.25	.43	.25	.49
Six-dimension	.17	.45	.17	.35

[4] Levine, *loc. cit.*

[5] Although a variety of other memory assumptions could have been made (see L. W. Gregg and H. A. Simon, Process models and stochastic theories of simple concept formation, *J. math. Psychol.*, 4, 1967, 246–276), these assumptions produce the highest predicted proportions of *S*s testing along the PD.

TABLE II

PREDICTED AND OBTAINED PROPORTIONS FOR TESTING ALONG SAME DIMENSION ON BLOCKS 1, 2, 3, AND 4: FOUR-VALUE TASKS

	Predicted	Obtained
Three-dimension	.037	.183
Four-dimension	.016	.155
Six-dimension	.005	.098

(Trials 16–20), *i.e.* completely eliminating all the values of their PD, following errors on Blocks 1, 2, and 3. The predicted proportions assumed that *S*s either have complete memory of information from the previous outcome trials and do not test disconfirmed Hs, or have no memory whatsoever of previous information and randomly select H from the set of tenable hypotheses. As can be seen, under all conditions *S*s tested along their PD to a greater extent than expected by chance. The differences between the obtained and predicted proportions were statistically significant in all cases.

Table III lists the predicted and obtained proportions of *S*s testing on Block $n+1$ an H that lay along the dimension of their H from Block n (this is analogous to a PH). The predicted proportions assumed either no memory or complete memory of the previous trial. These data are analogous to the data from Table I, part of which is rewritten here for comparison. The data from Blocks 3–4 are from *S*s who made three consecutive errors on the first three outcome trials. As can be seen, the data from Blocks 2–3 and Blocks 3–4 are almost identical to the data from Blocks 1–2, providing strong support for the observation that the dimension organizes hypothesis-testing behavior. All of the differences between the obtained and predicted proportions were statistically significant.

Tables I, II, and III show that following an error on Trial n, *S*s continued to test along that dimension on Trial $n+1$. This is not illogical behavior in that a disconfirmation of one value of a dimension does not imply the disconfirmation of other values of that dimension. (Erickson implied that a dimension could always be logi-

TABLE III

PREDICTED AND OBTAINED PROPORTIONS OF TESTING ALONG SAME DIMENSION FROM BLOCK n TO BLOCK $n+1$: FOUR-VALUE TASKS

	Predicted	Obtained		
	Block n to Block $n+1$	Block 1 to Block 2	Block 2 to Block 3	Block 3 to Block 4
Three-dimension	.33	.48	.49	.50
Four-dimension	.25	.49	.43	.56
Six-dimension	.17	.35	.29	.33

cally rejected following a disconfirmation of H.)[6] However, in a sequence of trials in which H was correct on Outcome Trial *n*, because it appeared with the correct value on that trial, and incorrect on Outcome Trial *n*+1, it would be illogical to test another H that lay along the Trials *n* and *n*+1 dimension. A dimension which has been inconsistently reinforced cannot logically contain the correct H.

Table IV compares the performance of *Ss* for whom dimensions were inconsistently reinforced (Column WRW) with those for whom dimensions were not inconsistently reinforced (Column WWW). Column WRW refers to *Ss* who made errors on the first and third outcomes, made a correct choice on the second outcome, and maintained the same H on Blocks 2 and 3. Column WWW refers to *Ss* who made three consecutive errors on the first three outcome trials. The entries in the table are the obtained proportions of *Ss* testing along the same dimension on Blocks 2 and 4 (not 2, *3*, and 4). As can be seen, the proportions under Column WWW are very similar and much higher than those under Column WRW. This indicates that Ss who have received inconsistent reinforcement on a given dimension are much less likely to continue testing along that dimension than are *Ss* who have not received inconsistent reinforcement. An omnibus test on the obtained proportions[7] led to the rejection of the statistical hypothesis that all the entries in Table IV are equal.

Table V presents data which deal with the question of selective memory loss following an error. In order to avoid confounding, only the data relating to the first outcome trial (Trial 5) were considered. In the table, the obtained probabilities are equal to the obtained proportions. The predicted probabilities assumed that Ss have no memory of the information received on Trial 5 and randomly select H. For example, for three-dimension, two-value tasks, there are six simple Hs; so it was predicted that following an error on Trial 5, the probability of *Ss* repeating the PH on Trials 6–10 would be

TABLE IV

Effect of Inconsistent Reinforcement along a Dimension: Four-Value Tasks

	WRW	WWW
Three-dimension	.13	.36
Four-dimension	.06	.33
Six-dimension	.07	.30

[6] J. R. Erickson, Hypothesis sampling in concept identification, *J. exp. Psychol.*, 76, 1968, 12–18.

[7] J. Cohen, An alternative to Marascuillo's "large sample multiple comparisons" for proportions, *Psychol. Bull.*, 67, 1967, 199–201.

.17 [Column P(PH)]. Furthermore, the predicted conditional probability of repeating the PH on Trials 6–10, given that *S*s chose within the PD, would be .50 [Column P(PH/PD)]. The third and fourth columns [P(SH) and P(SH_i/SD_i) respectively] present similar information regarding the SHs. For example, in a three-dimension, four-value task, there are two SHs on Trial 5 and a total of twelve Hs available for testing on Trials 6–10. Thus, the probability of testing one of the SHs on Trials 6–10, assuming no memory on Trial 5, would be .17.

As can be seen from Table V, all of the obtained probabilities were lower than the predicted probabilities; and with few exceptions, all of these differences were statistically significant. This indicates that for both PHs and SHs, *S*s have memory of previous outcome information. Of primary interest here is the differential memory loss for PHs and SHs. Observation of Columns P(PH/PD) and P(SH_i/SD_i) indicates that for both two-value and four-value tasks the probability of repeating a disconfirmed PH on Trials 6–10 was less than that of repeating a disconfirmed SH. This difference is not systematically related to number of dimensions. Statistical analyses confirm these observations.[8] For both two-value and four-value tasks respectively, the separate omnibus tests of equality of proportions were rejected at the .01 level of confidence. The only subsequent simple contrasts which were significant involved SH vs. PH comparisons.

Discussion

The data provide strong support for the Fishbein *et al.* observation that in concept-identification tasks, hypothesis testing is organized by the stimulus dimensions. The extent of this behavior does not appear to be systematically related to either number of dimensions or number of values within a dimension (Table I). Tables II and III indicate that this behavior persists beyond the first outcome trial and until all the values that lie along a dimension are exhausted. The data of Table IV suggest that testing along dimensions is a rational approach to learning the correct concept, in that *S*s showed great shifts away from an inconsistently reinforced dimension but continued to test along a dimension which was not inconsistently reinforced.

A fruitful framework within which to place this behavior is to consider it as a strategy.[9] As Table V shows, *S*s utilizing this strategy

[8] Cohen, *loc. cit.*

[9] J. S. Bruner, J. J. Goodnow, and G. A. Austin, *A Study in Thinking*, 1956.

TABLE V

Predicted and Obtained Probabilities for Testing PHs and SHs

	P(PH)		P(PH/PD)		P(SH)		$P(SH_i/SD_i)$	
	Predicted	Obtained	Predicted	Obtained	Predicted	Obtained	Predicted	Obtained
Three-dimension, two-value tasks	.17	.00	.50	.00	.33	.15	.50	.26
Four-dimension, two-value tasks	.12	.01	.50	.02	.38	.23	.50	.40
Six-dimension, two-value tasks	.08	.03	.50	.05	.42	.13	.50	.21
Three-dimension, four-value tasks	.08	.05*	.25	.10	.17	.10	.25	.19*
Four-dimension, four-value tasks	.07	.00	.25	.00	.18	.09	.25	.18*
Six-dimension, four-value tasks	.04	.02*	.25	.07	.21	.13	.25	.19*

* $p > .05$ for the difference between the obtained and predicted probabilities.

were much less likely to forget previous information than when they did not test along dimensions; *i.e.* the conditional probabilities of repeating disconfirmed PHs were less than the conditional probabilities of repeating disconfirmed SHs. This observation supports the suggestion of Fishbein *et al.* that testing along dimensions reduces *S*'s memory load. An alternative explanation consistent with the observed phenomena is that *S*s have dimensional preferences which determine the nonrandom hypothesis-testing along dimensions.

The data in Table V confirm our observation from Levine's data[10] that there is differential memory loss for PHs and SHs. This phenomenon was not systematically related to number of dimensions. As indicated above, the superior retention of PHs may reflect the operation of the superior strategy of testing along dimensions as opposed to other strategies. Alternatively, it is possible that this phenomenon is an artifact of the blank-trials technique. With this technique, *S*s are required to rehearse the PH through four nonreinforced trials and the following outcome trial, whereas the SHs can only be rehearsed on the outcome trial. It can be argued that the differences in rehearsal produced differences in memory. Unfortunately, there does not appear to be an obvious way in which to experimentally oppose these explanations. If the phenomenon turns out to be artifactual, then some of the above conclusions may need revision.

Summary

Adult human *S*s were run in either three-, four-, or six-dimension, two- or four-value concept-identification tasks. Using the blank-trials technique of Levine, it was found that *S*s learning the correct concept tended to eliminate all the values of a given dimension prior to testing along another dimension. This behavior was not systematically related either to number of dimensions or to number of values within a dimension. The other major finding was that *S*s had greater memory for the hypothesis held on blank trials than for the other hypotheses (values) which could have been tested on the outcome trials.

[10] Levine, *loc. cit.*

INPUT VARIABLES AND OUTPUT STRATEGIES IN FREE RECALL OF CATEGORIZED LISTS

By George Mandler, University of California, San Diego

Categorization and seriation have been proposed as two of the basic processes in the cognitive organization of word lists.[1] The phenomenon of categorization, or clustering, is well known; *i.e.* *Ss* learn categorized lists better than random word lists, and categories cluster in the output protocols for free-recall lists.[2] Seriation, the consistent serial ordering of items during output, has been shown to be the preferred method of organizing random lists in free recall when memory and processing mechanisms are not unduly overloaded. Mandler and Dean have shown that when lists of words are presented in an incremental manner, *i.e.* when a new item is added to a list on each trial, *Ss* will adopt for their output order the serial order of input whenever possible.[3]

In the present experiments both serial and categorical input conditions were presented to *Ss* in single lists to determine the extent to which these input variables affect the output strategies of *Ss*. In incrementing lists, we have shown that seriation is the preferred method, while categorized lists offer *Ss* the best opportunity for categorical organization. For nonincrementing, traditional presentation, clustering and categorization have been known to occur frequently, and Murdock has suggested that free, as against ordered, recall is the preferred method of recalling such lists of words.[4] Whether the additional cue of constant serial order affects the recall and organization of such organized lists is still an open question. The following experiments investigated the recall and organization

* Received for publication May, 1969. The research reported was supported in part by a grant from the National Science Foundation, GB-7807. The author wishes to express his appreciation for technical and editorial help to Peter J. Dean and Zena Pearlstone.

[1] George Mandler, Words, lists, and categories: An experimental view of organized memory, in J. L. Cowan (ed.), *Studies in Thought and Language*, 1969.

[2] W. A. Bousfield, The occurrence of clustering in the recall of randomly arrayed associates, *J. gen. Psychol.*, 49, 1953, 229–240.

[3] G. Mandler and P. J. Dean, Seriation: The development of serial order in free recall, *J. exp. Psychol.*, 81, 1969, 207–215.

[4] B. B. Murdock, Jr., The serial position effect of free recall, *J. exp. Psychol.*, 64, 1962, 482–488.

of precategorized lists under the incrementing method and under the traditional method with random and constant-serial input orders.

Method

General design. Three groups of *Ss* were used. One group learned a list of 16 words in an incremental manner with one new word, and only that word, presented on each trial (Group 1+1). A second group learned the same list of words but all words were presented on each trial in a new random order (Group 16×16R). The third group learned the same list with all words presented on each trial in an initial random order which remained constant from trial to trial (Group 16×16S). Each group was given two different lists with 17 trials each.

Subjects. A total of 36 *Ss* was used, with 16 *Ss* in Group 1+1 and 10 *Ss* in each of the other groups. All *Ss* were undergraduates fulfilling an introductory-course requirement.

Materials. Two lists were constructed consisting of 16 words each. Each list was constructed from the normative data of Cohen *et al.*[5] and contained the most frequently occurring words in each of four categories. The categories for one list were Furniture, Animals, Parts of Body, Methods of Transportation; for the other they were Fish, Kinds of Cloth, Professions, Musical Instruments. Half of the *Ss* in each group learned one of the lists and then were given the same task with the other list; for the other *Ss* the order of the two different lists was reversed. All word orders were arranged with the constraint that no item from any one category ever follow another item from that category. Five different random orders of the words were used.

Procedure. *Ss* were given full instructions on the design of the study and what to expect. Specifically, they were told that they were to learn a list of 16 words, each of which belonged to one of four categories which were then named for them. Words were presented on a Stowe memory drum at a 1-sec. rate. Following each trial, *Ss* were told to recall words in any order they chose, and they were given a sheet of paper on which to recall all the words available to them. The next trial was presented when *S* indicated that he had finished; however, in no case was *S* given more than 1 min. for recall. Following the acquisition of the first list, *E* changed tapes on the memory drum and *S* was given a new list in the same condition as List 1. Thus List 2 provides data on fully instructed, practiced *Ss*. In *Condition 1+1*, each trial consisted of the presentation of a single, new, different word. Thus, on Trial *n*, *S* was given the *n*th word in the list and only that word. On each trial *S* was required to recall all the words he had seen on that and all previous trials. No new word was presented on Trial 17; *S* was required only to recall the list again. In *Condition 16×16R*, the full list of 16 words was presented on each trial in a new random order. In *Condition 16×16S*, the full list of 16 words was presented on each trial in the identical, random order.

Data analysis. As a measure of organization the ITR(2) measure was used.[6] It is based on the number of repetitions of pairs of items in two successive

[5] B. H. Cohen, W. A. Bousfield, and G. A. Whitmarsh, Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, and Office of Naval Research.

[6] Mandler and Dean, *loc. cit.*

events, either an input and an output or two outputs. It is a bidirectional measure giving the ratio of observed intertrial repetitions (ITRs) to the maximum possible ITRs for the particular pair of events.[7] Expected chance values for ITR(2) are shown in Figs. 3 and 4, together with the obtained values. The second measure of organization was the relative ratio of repetition (RRR), which measures clustering with reference to the *E*-designated categories.[7] It is the ratio of the actual proportion of repetitions [Total repetitions (R)/ Words recalled (N)−1], over the maximum possible RR [Number of words recalled (N) − Number of categories recalled (N_c)/Words recalled (N) − 1], which reduces to RRR = R/(N − N_c).

Results

Performance. Fig. 1 shows the mean recall for Group 1+1 for Lists 1 and 2. The 16 *S*s in this group were divided into two groups, those who organized the list serially (the seriators) and those who organized

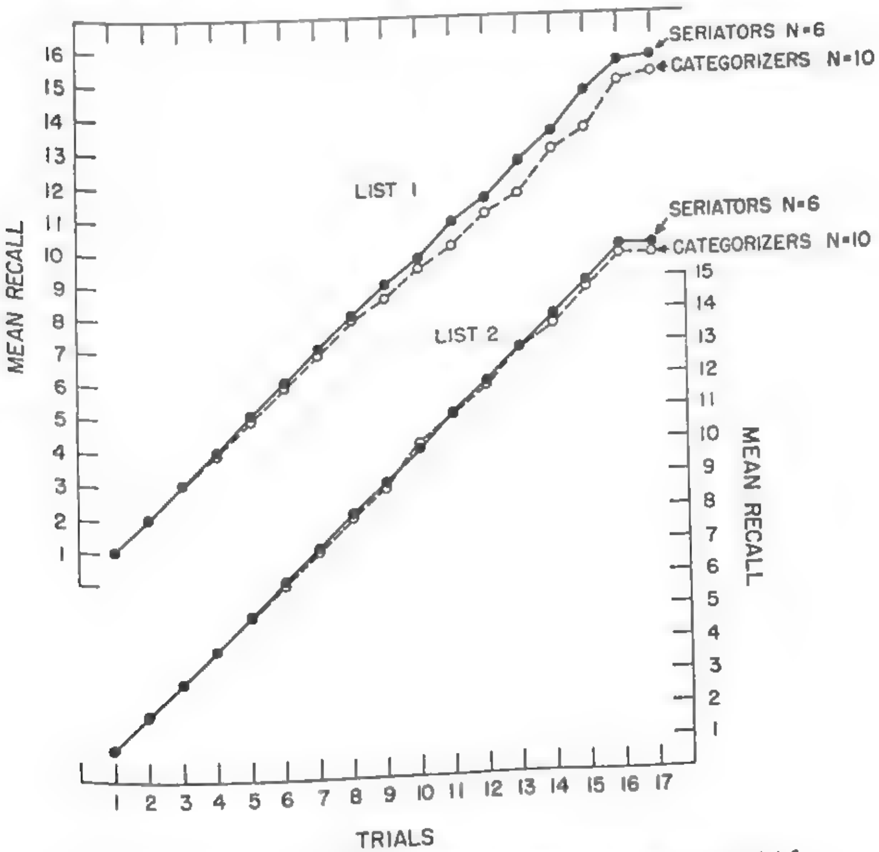


Fig. 1. Number of Items Correctly Recalled for Group 1+1, Lists 1 and 2, Segregated by Seriating and Categorizing *S*s

[7] A. K. Bousfield and W. A. Bousfield, Measurement of clustering and of sequential constancies in repeated free recall, *Psychol. Rep.*, 19, 1966, 935–942.

ized the list according to category clusters (the categorizers).[8] The figure shows that performance was at the expected high level for all groups, with the seriators slightly better than the categorizers but with all *Ss* giving essentially perfect performance on List 2.

Fig. 2 shows the performance for Groups 16×16S and 16×16R. The groups from the present experiment are indicated by the letters CAT, and for comparison are shown two groups that Mandler and Dean[9] ran under identical conditions except that the 16 word lists were random sets of 'unrelated' words. For both naïve and practiced *Ss*, performance was better for the categorized lists, as expected, and and for List 2 the superiority for the serial list was shown for both categorized and unrelated lists. Actually, Group 16×16 CAT S

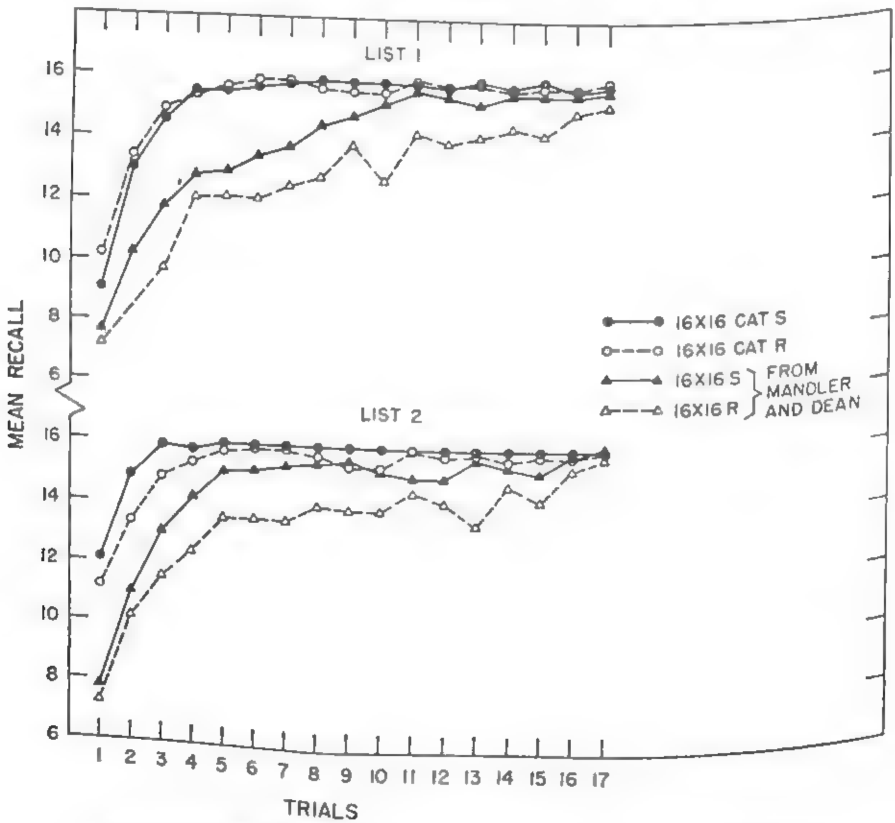


FIG. 2. MEAN CORRECT RECALL AS A FUNCTION OF 17 TRIALS FOR GROUPS 16×16R AND 16×16S
(Groups from the present experiment shown with the letters CAT; comparable groups from the study by Mandler and Dean shown without such letters)

[8] The criterion for this partition is discussed below.
[9] Mandler and Dean, *loc. cit.*

reached essentially perfect performance by Trial 3. While the two sets of data were run for different studies, *Ss* were obtained from similar subject pools and sampling differences are unlikely to account for the large differences.

Organization. Fig. 3 shows the ITR(2) measure for Group 1+1, for both the Input-Output (IO) and the Output-Output (OO) relations for both lists. The distinction between seriators and categorizers was made by a simple inspection of their individual ITR(2) scores. When the IO values were high, *Ss* were seriators, *i.e.* their output was organized in accordance with input order. The serial order of the appearance of the words was considered as the input

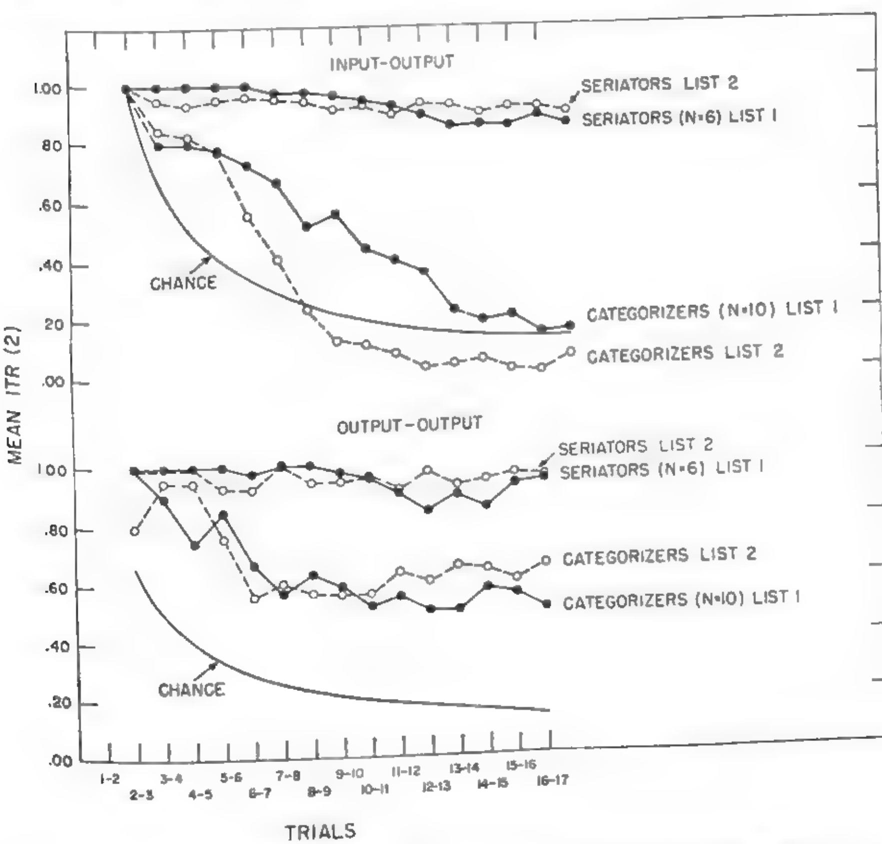


FIG. 3. MEASURE OF INTERTRIAL REPETITION [ITR(2)] FOR GROUP 1+1 OVER 17 TRIALS FOR LISTS 1 AND 2

(Continuous 'chance' line shows mean expected chance value computed on the basis of actual number of words recalled; Top graph shows Input-Output [IO] concordance; lower graph that for Output-Output [OO]; for IO relation, measure computed for successive trials, comparing input on that trial with output; for OO, successive measures computed for output on trials n and $n+1$, as shown on baseline)

order of the items. This classification of *S*s was unequivocal since no *S*s fell into an indeterminate category. The IO graph shows the result of this selection: seriators had a mean ITR(2) value around .90, while categorizers, who presumably organized the lists according to categories and clusters, showed an ITR(2) value near chance. The below-chance value for the practiced categorizing *S*s could have been expected since the input lists were arranged in such a way that members of *different* categories always occurred contiguously in the input order; chance values would expect such contiguities to occur in output as well. When *S*s categorize perfectly, their output would show very few such occurrences (ideally, only three), and therefore their IO values would be below chance. The OO values show significant organization for both subgroups, with the seriators showing the expected higher organization values, since their output was determined primarily by the input order which was nonvariable.

Given that the seriators organized the list in accordance with the word order of the input list and that categorizers presumably used the *E*-determined categorical structure, the RRR measure gives further evidence on this point. Table I shows the RRR values for Trials 16 and 17 for Lists 1 and 2 for the two types of *S*s and includes expected chance values. The evidence clearly shows high levels of clustering—practically perfect for List 2—for the categorizers and no clustering, of course, for the seriators. The below-chance value for the seriating *S*s could have been expected since the input lists were arranged in such a way that no two members of the same category ever occurred contiguously in the input order and, therefore, would not appear contiguously in the output of seriators. Chance values, on the other hand, would expect some clustering of members of the same category by chance alone.

Fig. 4 shows the ITR(2) values for the two 16×16 groups. It is evident that none of the 20 *S*s in these two groups seriated, since their IO data are consistently at or below chance level. The OO data show clearly superior organization on the part of the *S*s who were

TABLE I

MEAN RRR SCORES FOR LAST TWO TRIALS OF TWO SUBGROUPS IN CONDITION 1+1

	List 1		List 2	
	Trial 16	Trial 17	Trial 16	Trial 17
Seriators (N = 6)	.11	.09	.04	.04
Categorizers (N = 10)	.71	.75	.98	.97
Expected chance value	.20	.20	.20	.20

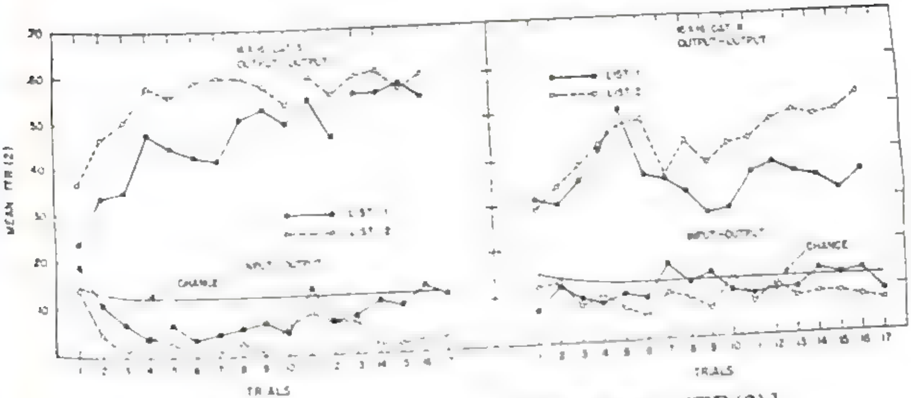


FIG. 4. MEASURE OF INTERTRIAL REPETITION [ITR(2)] FOR GROUPS 16×16R AND 16×16S OVER 17 TRIALS FOR LISTS 1 AND 2

(Top graphs show Output-Output [OO] concordance; lower graphs that for Input-Output [IO]; continuous 'chance' lines show mean expected chance value computed on basis of actual number of words recalled)

given constant input lists (16×16S) and improvement in organization for both groups from List 1 to List 2.

The superior organization of the categories for Group 16×16S is also indicated by the RRR measure in Fig. 5. The constant-input *S*s showed higher clustering than the random-input group by the later trials; $F\ (16, 288) = 1.56, p \approx .08$. Chance values for these data were .18 for Trial 1 and .20 for all other trials.

DISCUSSION

The performance data generally confirm previous findings. Incrementing lists, even when words are presented only once, show a high level of acquisition and performance. The output strategies of *S*s appear to make little difference; however, the apparent ceiling effect in these data does not permit any definite conclusion in this respect. The traditional method of presenting lists shows the expected superiority of categorized over 'unrelated' lists, with some remaining apparent advantage for constant serial input; but again it is impossible to reach any definite conclusion because of ceiling effects.

The organizational data show that when memory and processing capacities are not overloaded (*i.e.* with incremental presentation) and when *S*s are aware of the categorical structure of the list, some *S*s choose categorization, while others prefer to seriate the list. This distinction between seriators and categorizers as an individual-difference variable may be of future use in defining preferred general

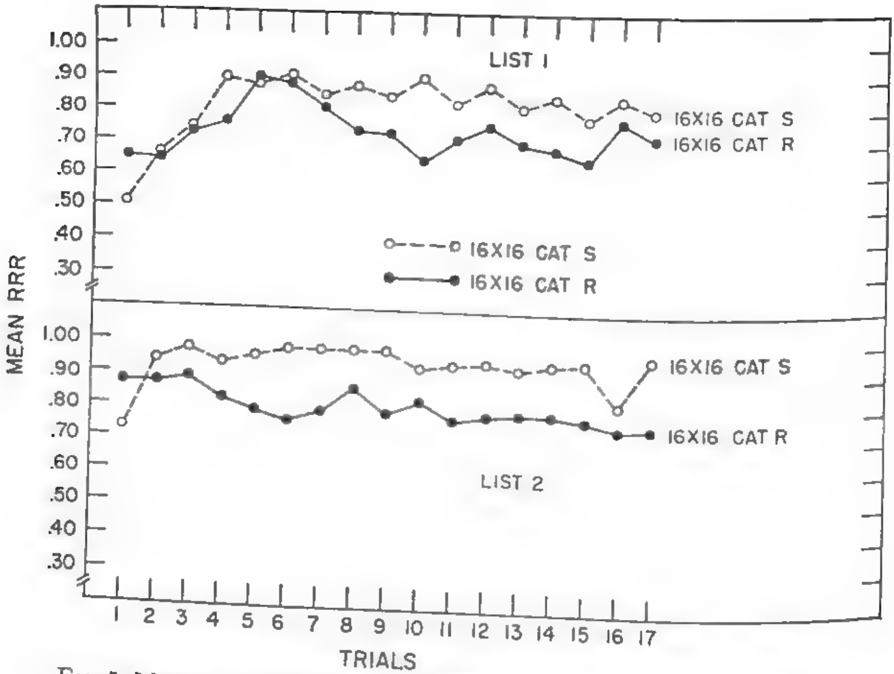


FIG. 5. MEAN RELATIVE RATIO OF REPETITION (RRR) FOR 17 TRIALS FOR GROUPS 16×16R AND 16×16S ON LISTS 1 AND 2

strategies in individuals, but for the present time we can only conclude that both of the major modes of organization may be used. It is interesting to note that the *S*s who seriated or categorized List 1 also used the same strategy on List 2. The traditional presentation of lists produces consistent categorization by all *S*s. This would, of course, be expected for the random-input group, since the input condition does not provide *S*s with consistent serial cues to be used during output. However, the serial group ignored the constant input order and categorized the list, actually achieving the higher level of clustering performance.

Mandler and Dean have suggested that seriation is the preferred mode of organizing lists when memory and processing capacities are not overloaded.[10] This conclusion must be modified by noting that when categorical structure is obvious and easily adopted, some *S*s use categorization as their preferred strategy in organizing output. In the traditional input presentation, seriation tends not to be used with categorized lists, and here Murdock's conclusion that free recall (*i.e.* organized recall) is preferred to serial recall[11] holds.

[10] Mandler and Dean, *loc. cit.*
[11] Murdock, *loc. cit.*

The superiority of the constant-serial-input list over the random-input list in clustering and organization deserves some additional attention. One argument might be made that constant serial input provides *Ss* with additional cues so that the members of specific categories may be easily located during input and thus generate better organization. A more attractive position suggests that the random input order provides an interfering and confusing element. Thus the appearance of items in input next to various different items on successive trials may produce inadvertent clustering which is adopted and discarded until the major categories are finally adopted as major retrieval clues. The RRR data suggest some support for that position, in that the clustering score of the constant-input group was similar to that of categorizing *Ss* in the incrementing group. For Group 1+1 no argument could be made that the input condition provided additional position cues for the location of the items in the input list, since there was no repeated-input list. The similarity of the RRR scores for Groups 16×16S and 1+1 suggests that the lower RRR value of Group 16×16R was due to interference in the random-input condition. This suggests another defect of the traditional random-input methods for free-recall experiments: they not only overload the memory and processing capacities of *Ss* but also interfere with optimal organization of items in categorized lists.

Summary

The *Ss* were required to learn categorized lists under three conditions: (1) under incrementing conditions, when only one new word was presented on each trial, (2) with constant serial order of the input list, and (3) with input order randomized from trial to trial. The first condition, which minimizes memory and processing overload, produced seriation or categorization as output strategies. Organization and clustering were superior for the constant-serial-input list over the random-input lists, even though all *Ss* in these conditions adopted a categorization strategy. The suggestion was made that random-input lists interfere with optimal organization of categorized lists in free recall.

THE ROLE OF RHYTHM IN ACTIVE VERBAL MEMORY: SERIAL INTRUSIONS

By Ulric Neisser, Cornell University

Subjects in memory-span experiments nearly always impose their own rhythmic organization on the strings of digits they are to remember. Presented with an unpatterned stimulus, say, *4186037295*, they are likely to reply '418–603–729–5' or with some other grouped pattern. Why do they do this?

One hypothesis, which I have put forward elsewhere,[1] is that the active verbal memory which underlies such performances cannot exist at all except as an organized structure and that rhythm is the simplest form of organization available to *S*. He would be better off if he could relate the stimuli to some kind of mathematical progression, or to a series of historical dates, or (if he had been given a string of words) to the syntactic structures of his language; but lacking any of these, he can at least organize the verbal responses themselves into temporal groups. It is because this grouping pattern exists that *S* can rehearse the digits internally until response is required; it is through the grouping pattern that he can assign positions to particular digits, and often recover a digit later in the sequence even when he has forgotten an earlier one.

This hypothesis suggests that the digit span should be drastically *shortened* when the rate of presentation is so high that rhythmic organization is impossible, and substantially *lengthened*, at normal rates, if *S* manages to learn more sophisticated grouping techniques. Both of these predictions have been confirmed in the literature.[2]

* Received for publication January 23, 1969. The substance of these results was reported at the meetings of the Eastern Psychological Association in April, 1968. The research was conducted at the Unit for Experimental Psychiatry, affiliated with the Pennsylvania Hospital and the University of Pennsylvania. It was supported in part by Contract No. Nonr-4371(00) from the Office of Naval Research and in part by the National Institute of Mental Health through a Research Career Award to the author. The assistance of Mr. William Orchard, who ran some of the *Ss*, is gratefully acknowledged.

[1] U. Neisser, *Cognitive Psychology*, 1967, 219–242; U. Neisser, Yvette J. Hoenig, and Evelyn Goldstein, Perceptual organization in the prefix effect, *J. verb. Learn. verb. Behav.*, 8, 1969, 424–429.

[2] D. B. Yntema, F. T. Wozencraft, and L. Klem, Immediate recall of digits presented at very high speeds, a paper presented to the Psychonomic Society, 1964; P. R. Martin and S. W. Fernberger, Improvement in memory span, this Journal, 41, 1929, 91–94.

The present experiment tested a different kind of prediction: one which does not deal with the number of digits recalled correctly but with wrong responses. The particular class of errors to be studied was first reported by Conrad, who called them "serial order intrusions."[3] Such an intrusion occurs when the nth digit of the response seems to come from the nth member of the *preceding* string instead of repeating the nth digit of the string just presented. Some examples appear in Table II: both the "8" in the second response and the "9" in the third response are serial intrusions.

On the hypothesis put forward here, the serial positions of digits in such a string are defined for *S* only by the rhythmic organization he imposes on it. If this is true, genuine intrusions should occur only when successive lists have the *same rhythmic pattern*. When a string has a different rhythm from the preceding one, their nth positions should be independent, and the same digit should not appear in both except by chance. In the main experiment reported below, the sequence of grouping patterns was arranged to test this deduction. In half the trials, the string of digits had the same grouping pattern as the one that had just preceded it; in the other half, this was not the case. For reasons explained below, a control experiment was also conducted. In the control experiment, all strings had the same grouping pattern, but some were presented to *S*'s left ear and some to his right.

Method

Stimulus materials. For the main experiment, 113 strings of 10 digits were prerecorded binaurally on magnetic tape. Each was a random permutation of the digits from 0 through 9, except that the nth digit of a string was never identical with the nth digit of the string before. Natural runs of more than two digits (like *456*) were eliminated. The digits were spoken at a rate of 2 per sec., except that additional pauses of .5 sec. were introduced to create certain grouping patterns. The first three strings were used for practice only. The remaining 110 were arranged in 10 sets, each set containing 11 strings of digits. The interval between the strings within a set was 10 sec. Between sets, *S* could rest if he wanted to.

Each string of digits was grouped in one of two ways: a *triadic,* or 3–3–3–1 pattern, and a *quartic,* or 4–4–2. They were arranged in double alternation within each set, as shown in Table I, which illustrates the first of the 10 sets used. If the first trial is ignored (since it cannot suffer any serial intrusions), such a set contains five strings that have the same rhythmic structure as the one before, and five others with a different pattern than their predecessor. These may be termed *congruent* and *mismatched* respectively. Each *S* was

[3] R. Conrad, Serial order intrusions in immediate memory, *Brit. J. Psychol.*, 51, 1960, 45–58.

thus presented with a total of 50 congruent and 50 mismatched strings. The experimental hypothesis was that serial intrusions would occur beyond the chance level only in the congruent ones.

Subjects. The main experiment employed 19 *Ss*, 16 men and 3 women. All were college students who had volunteered and were paid for their services.

Procedure. The *Ss* listened to the stimuli over earphones. They were asked to repeat back each string of digits as soon as it was over. (It was felt that written response might have reduced *S*'s tendency to use rhythmic grouping as an organizing factor.) *E* recorded the responses by marking a data sheet on which the correct strings were listed.[4] The 10-sec. interval between successive strings proved adequate for this purpose. The *Ss* were not informed of the aim of the experiment, nor of the special features of the experimental design. However, all of them noticed the variations of grouping in the stimulus strings.

Results

The variable of interest is the proportion of errors which were serial intrusions. Since there were nine possible errors at any point (any of the 10 digits except the correct one), the chance rate of serial intrusions was 1/9. This fraction represents the null hypothesis that successive strings are independent. However, it cannot be directly applied to all the errors; it applies only to those where serial intrusions could conceivably have occurred. These were defined here by two criteria: first, *S* must have made an identifiable wrong response (not just an omission); second, he must have made a *correct* response to the corresponding digit of the preceding string. It seemed unwise to expect much intrusive tendency from a digit which *S* had missed, and perhaps not heard at all, or from a response that may have represented a mere guess. Table II clarifies the definition used.

Table I

First Set of Stimulus Strings

7 9 4 – 0 5 1 – 6 3 2 – 8
8 2 9 – 3 6 4 – 0 1 5 – 7
6 5 7 1 – 8 0 3 9 – 4 2
7 1 0 9 – 5 2 8 4 – 3 6
2 4 5 – 2 7 6 – 0 8 9 – 1
9 3 6 – 0 4 2 – 5 7 1 – 8
8 5 1 3 – 6 9 0 2 – 7 4
3 7 9 4 – 1 6 5 8 – 0 2
1 8 3 – 9 4 2 – 7 0 5 – 6
6 2 0 – 8 3 5 – 9 1 4 – 7
1 3 7 5 – 9 6 2 8 – 0 4

[4] When *S* indicated by a pause or an explicit remark that he was omitting one or more digits, *E* recorded blanks. Responses following a blank were scored only if they were in convincing correspondence with particular stimulus items, usually because of a falling intonation at the end of the string. About 8% of the response strings had internal blanks created in this way. A separate data analysis carried out without these strings gave substantially the same results as those reported here.

TABLE II

EXAMPLES OF INTRUSIONS

String No. 1	*4*	*1*	*8*	*6*	*0*	*3*	*7*	*2*	*9*	*5*
Response	"4	1	8	0	6	3	7	9	·	·"
String No. 2	*5*	*6*	*1*†	*9*	*4*	*7*†	*8*†	*0*	*3*	*2*
Response	"5	6	8‡	9	4	2	6	3	0	2"
String No. 3	*7*	*4*	*3*	*8*†	*6*†	*0*	*2*	*5*	*9*	*1*
Response	"7	4	3	9‡	5	7	·	6	0	1"

† Opportunities for intrusion.
‡ Intrusions.

Note that only three of the errors in the second string and two in the third offered 'opportunities for intrusion.' As it turned out, about 48% of the errors that actually occurred in the experiment were opportunities for intrusion by these criteria. The 'intrusion ratio' was defined as the number of intrusions divided by the number of opportunities.

The intrusion ratios obtained are displayed in Fig. 1, which shows the four frequency distributions. The upper two are for the main experiment, and each is based on 18 *Ss*. (One *S* was discarded because she made virtually no errors, so no intrusion ratio could be calculated for her.) On the null hypothesis, all these ratios should average around .11. This was the case for the series mismatched in rhythm (first histogram). While one *S* did have a ratio above .17, and one fell below .05, the mean for mismatched series was slightly below chance expectation at .095.

Very different results appeared when successive strings had the same rhythm, as shown in the second histogram. The mean was well above chance (at .179), and so were 15 of the 18 *Ss*, a highly significant result by a binominal test ($p < .01$). Six *Ss* had ratios which were significantly ($p < .05$) above .11 even when considered *individually*, by a binominal test against the null hypothesis with $p = 1/9$. Moreover, if each *S* is considered as his own control, 16 out of 18 had a higher intrusion ratio in the congruent than in the mismatched strings.

Two other results of the main experiment are worth noting, although they do not bear directly on the hypothesis. There was no significant difference between the numbers of *opportunities* for intrusion made in the two kinds of strings. The mean per *S* was 52.2 opportunities (out of 500 digits) in the congruent strings and 53.3 in the mismatched strings. This means that the additional intrusions in the congruent strings do not represent additional errors. Intrusions apparently represent instances where *S* has *already* forgotten the

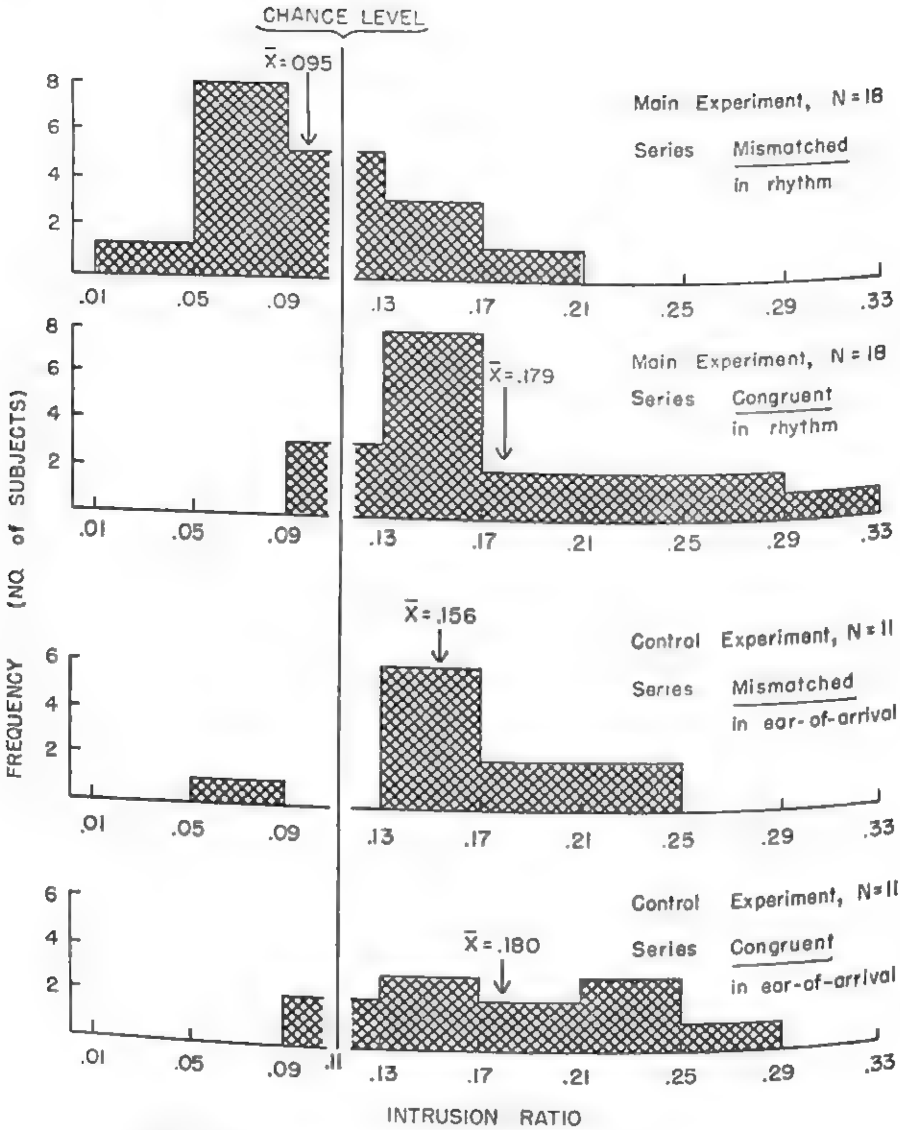


FIG. 1. RESULTS OF MAIN AND CONTROL EXPERIMENTS

correct digit for other reasons, and emits the intruding digit rather than guessing at random. This finding is not new; Conrad obtained a similar result, though he varied the time interval between successive strings rather than their congruence in rhythm.[5] The second minor result is that the mean number of errors (of all kinds) per *S* was 98.9 in the quartic (4–4–2) strings and 123.7 in the triadic (3–3–3–1) strings; the difference is significant at the .01 level by

[5] Conrad, *loc. cit.*

Wilcoxon's T test. Of the 18 *Ss*, 14 made more errors in triadic strings, 4 in quartic strings. A number of hypotheses might be advanced to explain the greater difficulty of the triadic strings (they take longer, contain more groups, etc.), but this experiment did not provide support for any particular one of them.

Control Experiment

The main experiment showed that extrachance serial intrusions occur only when successive strings have the same rhythm. However, it might still be argued that the important variable is not congruence in rhythm per se but simply the general similarity between the strings in question. Serial intrusions might be seen simply as instances of generalization, which is known to depend on the similarity of stimulus patterns.

In its broadest form this hypothesis cannot be refuted, but its plausibility can be examined by manipulating a *different* dimension of similarity. The dimension chosen was ear-of-arrival, *i.e.* the side on which the stimulus was presented. Differences in ear-of-arrival are very noticeable and play a major part in other cognitive phenomena.[6] Therefore, a simple generalization hpyothesis might predict that serial intrusions would be much reduced when successive strings were presented to opposite ears. However, the rhythm hypothesis predicts that such 'mismatching' should make *no* difference in the intrusion ratio.

Stimulus materials. The same series of 113 strings was recorded using triadic grouping throughout. The first three strings were presented binaurally, for practice. Half of the remainder were presented to *S*'s left ear, the other half to his right. The sequence of left and right ears followed the same double-alternation schema as the sequence of grouping patterns in the main experiment. Thus, half the strings were *congruent* and half were *mismatched* with respect to ear-of-arrival.

Subjects. Eleven additional *Ss* (eight men and three women) were recruited in the same way as before.

Procedure. The procedures of the main experiment were followed throughout, except for the difference in mode of stimulus presentation.

Results. The two lower histograms in Fig. 1 show the intrusion ratios obtained in the control experiment. As predicted, serial intrusions were *not* substantially reduced when successive strings arrived at different ears. A slight decrease did appear (the mean of

[6] See, for example: E. C. Cherry, Some experiments on the recognition of speech, with one and with two ears, *J. acoust. Soc. Amer.*, 25, 1953, 975–979; A. M. Treisman, Selective attention in man, *Brit. Med. Bull.*, 20, 1964, 12–16.

intrusion ratios was .180 in congruent series and .156 in mismatched series), but it was negligible compared with the shift produced by rhythmic mismatching in the main study. Ten of the 11 control *Ss* had intrusion ratios above $1/9$ even in the mismatched condition, with successive strings arriving from opposite sides.

Conclusions

The results of the main study showed that serial intrusions do not occur beyond the chance level when successive strings have different grouping patterns. No corresponding reduction in intrusions was produced by presenting successive strings to opposite ears. These findings are consistent with the hypothesis that rhythmic grouping is the structural basis of active verbal memory.[7]

Summary

Performance in memory-span tasks is assumed to depend on *S*'s elaboration of a rhythmic grouping structure, in which each digit has a place. This suggests that serial intrusions (responses which were appropriate to a given position in the preceding string of digits) should occur only if successive strings have the same grouping structure. An experiment with 18 *Ss* confirmed this hypothesis.

[7] Related findings concerning the role of grouping may be found in two articles which appeared after this paper was in press: J. Ryan, Grouping and short-term memory: Different means and patterns of grouping, *Quart. J. exp. Psychol.*, 21, 1969, 137–147; J. Ryan, Temporal grouping, rehearsal, and short-term memory, *Quart. J. exp. Psychol.*, 21, 1969, 148–155.

NOTES AND DISCUSSION

THE SIXTY-FIFTH ANNUAL MEETING OF THE SOCIETY OF EXPERIMENTAL PSYCHOLOGISTS

The sixty-fifth annual meeting of the Society of Experimental Psychologists was held at Mount Holyoke College, South Hadley, Massachusetts, on April 4 and 5, 1969. John Volkmann, Chairman of the Society for the year, presided at the business meeting and at sessions for the presentation of scientific papers.

Twenty-two members and fellows were present: Blough, DeValois, Garner, Geldard, E. J. Gibson, J. J. Gibson, Green, Helson, Jenkins, Kamin, Kappauf, Kendler, Meyer, Newman, Pfaffman, Riggs, Solomon, Stevens, Teitelbaum, Underwood, Volkmann, and Wickens. Richard J. Herrnstein, Julian Hochberg, William Kessen, George Mandler, Endel Tulving, and Eliot S. Valenstein were elected to membership in the Society, bringing the total to 71 members and 39 fellows.

Reports of research were presented in scientific sessions on both days of the meeting. The Warren Medal for 1969 was awarded to Roger W. Sperry "for his insightful use of the split-brain method of investigating the roles of the cerebral hemispheres, both separately and in their interaction."

The Society accepted the invitation of the University of Pennsylvania to meet in Philadelphia, Pennsylvania, in 1970. Richard Solomon was elected Chairman for 1969–1970.

University of Illinois, Urbana

WILLIAM E. KAPPAUF
Secretary

BOOK REVIEWS

Edited by JULIAN HOCHBERG, Columbia University

VERBAL LEARNING. By JOHN JUNG. New York: Holt, Rinehart and Winston, 1968. Pp. ix, 212. $5.95.

Jung has written a text for undergraduates which attempts to survey the current experimental literature on verbal learning. This new book is organized by chapters which treat broad topics, *e.g.* acquisition, transfer of training; and these are further divided into subtopics, *e.g.* response availability, transfer paradigms. The method of presentation is unusual: for each area a few typical experiments are described and these serve as the base for the methodological and theoretical points the author wishes to make. The writing style, while not lively, is clear and direct, and it is effective when an experiment is described or when some point of method is made (and this comprises most of the book). The problems treated and their manner of treatment are what one would expect from a contemporary associationist, and the front of the book's jacket confirms this expectation. On it there is a human face (the face of *S*, I believe) with jaws, lips, and articulatory mechanisms prominent and active. The eye, receiving the stimulus, is shown, but significantly, only a small portion of the cranium is visible. It is fun to try to determine what the lips are saying. 'WAB' is my guess.

To say that the text is written within a certain tradition, however, is not to say that it is explicit about that tradition. The book is in fact, quite atheoretical. It tells the student a good deal about research being done but does not tell him the purpose of this research. For example, in discussing the role of repetition in associative learning, Jung states that associationists and cognitive theorists are divided on this issue. But why? What is it about their beliefs that leads them to take such different positions? The chapter on transfer of training is another example of what I have in mind. The bulk of the literature on this topic makes sense only if one assumes the theory of transfer of identical elements. The very method (A-B, A-C paradigms, and the like) is merely a reflection of the theory. Nowhere does the author make this simple point clear. Nowhere is there discussion of the merits of alternative ways in which one might try to formulate the problem of transfer, nor is there any consideration of transfer in any setting other than the familiar two-list situation. Rather, the focus is on questions like 'Why is there more negative transfer in the A-B, A-C paradigm than in the A-B, C-B paradigm?' With few exceptions the book is concerned with experiments whose purpose in any larger sense is left obscure.

The combination of an implicit theory and a failure to come to grips with broader issues greatly reduces the value of the book as an introduction to the study of human learning; implicit theories tend to become biases, and failure to relate experiments to broader questions tends to discourage interest in that

literature. I believe that the sterility of a good deal of the research in this field stems from this same combination of faults.

Institute for Cognitive Studies
Rutgers University, Newark
JOHN CERASO

ATTENTION IN DISCRIMINATION LEARNING. By ELIJAH LOVEJOY. San Francisco: Holden-Day, 1968. Pp. xii, 133. $8.50.

This brief monograph in hard covers is a slight modification and extension of the author's doctoral dissertation. It is a fairly standard stochastic model of discrimination learning combined with additional assumptions in order to account for recent data in the area of selective attention. The model assumes that the animal's choice on any given trial in a simultaneous discrimination is controlled by only one stimulus dimension even though others may be physically present. The probability that a given dimension will be the controlling one is a joint function of two variables. These are the relative distinctiveness or attention-getting quality of that dimension and the relative response strengths associated with the two stimulus values on that dimension. If the animal remembers which stimulus dimension controlled its choice on that trial (and according to the model, it need not do so), the outcome on that trial modifies both the distinctiveness and the response-strength variables. As the author indicates, this is an attempt to specify more exactly the trial-by-trial changes in behavior that occur in two-stage theories of discrimination learning. The complexity of these theories is indicated by the fact that even this simple version requires $3N + 6$ parameters to predict behavior in a discrimination involving N different stimulus dimensions. Consequently, the only support offered for this model comes from a limited number of computer-simulation studies using an arbitrary set of parameter values. The results suggest that the model can reproduce a number of the findings reported in the experimental literature. In summary, this is a small book with a large price tag.

Stanford University
DOUGLAS H. LAWRENCE

THE TECHNOLOGY OF TEACHING. By B. F. SKINNER. New York: Appleton-Century-Crofts, 1968. Pp. ix, 271. $2.95.

Any new book by Professor Skinner is bound to occasion a great deal of interest among several audiences, especially if it contains material that is new, such as alterations in his viewpoint, new research, or applications of his ideas to topics to which they have not heretofore been directed. The volume under review, a paperback in the Century Psychology Series, however, includes virtually none of these features. In it are eleven chapters, four of them previously published, three here published for the first time but read on earlier occasions, and four which were prepared for this book.

The book concerns teaching, and it contains various criticisms of conventional teaching methods and their failures, but its major task is to advocate changes in teaching methods in line with the principles of operant learning. These ideas are communicated, with some supporting data, in the reprinted chapters. Here, as elsewhere in the book, positive reinforcement is emphasized (but nowhere defined or carefully articulated, this being a book designed for a

general audience); and behavioral outcomes of teaching are stressed, as opposed to the vague and general goals often held forth as the desirable outcomes of instruction. The argument and the ideas are familiar ones, but it is valuable to have in one place many of Professor Skinner's ideas about teaching.

The previously unpublished essays deal, in a sense, with special topics, such as teaching thinking, motivation of students, and creativity. These topics have, of course, been treated elsewhere, in *Verbal Behavior* and in *Science and Human Behavior*, for example. The new chapters include one on the etymology of words dealing with teaching and learning, and three viewpoints popular in education. There is one which discusses the use of punishment (the evaluation being not totally negative), techniques of self-management, and techniques of dealing with respondent behavior. A chapter which "reviews" teaching discusses the character of knowledge and confronts the problem of how one can initiate the behavior that is to be taught. Policy matters, administration, and training teachers form the substance of the provocatively titled chapter "Behavior of the Establishment."

This is an interesting, well-written, sometimes amusing set of essays, and they present Skinner's viewpoint toward teaching well and attractively. The claims made here for the operant technology do not seem to me excessive, and the book seems to be pervaded by modesty and recognition of difficulties. One may well disagree with Professor Skinner on a number of points, as I do, for example, in the easy invocation of the term 'reinforcement' and in the brief account of learning to talk. But as a source for finding out what Skinner thinks about teaching this book can be recommended.

Pennsylvania State University CHARLES N. COFER

PSYCHOLOGICAL TIME IN HEALTH AND DISEASE. By JOHN COHEN. Springfield, Ill.: Charles C Thomas, 1967. Pp. xv, 103. $6.50.

This very slim volume presents classical time psychology to the general reader. As such it offers a well-struck balance between the anecdotal and experimental evidence, although it leaves out some of the more recent quantitative approaches to the problems of subjective time and duration.

In Chapter 1 we are introduced to the concept of the internal clock as a complex set of endogenous rhythms and events. Cohen takes a healthily critical point of view toward the Piéron-Hoagland chemical hypothesis, which appears to affect the higher mental activities rather less than has sometimes been claimed. Chapter 2 deals with the psychophysics of simultaneity, order, and short intervals. An interesting set of experiments on the time needed to estimate subjective probabilities of success is reported, but failed to strike me as relevant to the main issue of this chapter. Chapter 3 deals with the "time perspective": the appreciation of long intervals with respect to one's own history. Chapter 4 reports on the "relativity of subjective time" and the genesis of the time concept. Time is influenced by the events which occur in it, and vice versa. These effects—the latter of which was discovered by the author—are known as the tau and kappa effects respectively. It is regrettable that there is no mention of Mashour's excellent work on the interaction of time, velocity, and distance, because it more nearly unties this Gordian knot than does the ancient effort of J. F. Brown.

The contents of Chapter 5 are predominantly anecdotal. They deal with the quite colossal "relativity" of subjective time in mental disorder. How impossibly complicated theory would be if we were to take into account all the idiosyncratic effects discussed in this chapter, in a theory of time experience! Chapter 6 then extends the argument into the fields of culture and religion. Time usually appears as a negative, *i.e.* destroying or devouring, power (a feeling that occasionally will engulf students of time). Chapter 7 makes a step "towards a model of psychological time." This model, which essentially consists of a pulse-generating mechanism, a counter and a comparator for comparing successive counts (intervals), is extended to account for several of the effects mentioned in the preceding chapters. These extensions, however relevant, strengthened my conviction that instead of trying to describe the relations between time and information processing by a counting mechanism of the sort presented here, we could explain the very same effects directly in terms of the information actually processed. But this appears to be a matter of isomorphism, and consequently of taste.

If I am to raise any serious criticism, it is only with the appearance of the book. For a price that buys the average textbook, one would expect better draftsmanship, less overlap between the contents of diagrams and tables, a single list of references, a better subject index, and fewer blank pages.

JOHN A. MICHON
Institute for Perception, RVO-TNO
National Defense Research Organization, TNO
Soesterberg, the Netherlands

ON HUMAN COMMUNICATION: A REVIEW, A SURVEY, AND A CRITICISM. By COLIN CHERRY. 2nd ed. Cambridge: The M.I.T. Press, 1968. Pp. xiv, 337. $2.95.

The second edition of this book has undergone few changes from the widely known previous edition, which was reviewed in this journal (G. A. Miller, 70, 1957, 332–333). This book still represents a well-conceived and brilliantly written survey of many of the disciplines and points of view which contribute to an understanding of human communication. It may be the only place where the reader can be introduced to such diverse concepts and techniques as semantic information, pattern recognition, Bayes's theorem, spectral analysis, and signal-detection theory. The book's broad outlook serves as a valuable correction to the intense compartmentalization found within experimental psychology.

The last decade of research on communication is scarcely represented at all in the book. Of the 43 new references, only 20 are from the 1960s and even these are used only for very minor modifications. Of course, a book for the general reader could not be expected to represent the latest developments in the individual fields, but the new edition does not do for current work in communication what the older edition did in its time. For example, there is no reference to the linguistic work of Chomsky, certainly a key to much current thinking about communication. One feels that whereas a book on communication written in 1956 could reasonably concentrate on the recent and exciting statistical and engineering developments, a book of the 1960s really needs to do more for the linguistic and psychological aspects. Nonetheless, any serious

student of the psychology of communication would still do well to start with Cherry. It is perhaps comforting that, today, there are more books to move on to after finishing it.

University of Oregon MICHAEL POSNER

LOGICAL THINKING IN CHILDREN. Edited by IRVING E. SIGEL and F. H. HOOPER. New York: Holt, Rinehart and Winston. 1968. Pp. xv, 541. $7.95.

The title is misleading; this is an edited collection of papers dealing not with logic in the general sense, nor with formal operations in the narrow sense, but with concrete operations thinking. The editors have focused upon this developmental stage for a very good reason: the bulk of available research evidence is concerned with it. The book is very sensibly organized into six chapters, each with a preface which introduces the basic concepts and briefly integrates the selections included, concluding with a list of questions for future research. The six substantive sections deal with: conservation of quantity, number concepts, spatial and geometric concepts, logical operations, training research, and theoretical discussions. A final summary chapter by Sigel deals with related issues not specifically covered in the first six sections. Each of the selections included meets with American methodological preferences with respect to standardization of procedure, specification of *S* population, and use of statistics; in addition, each has enough theoretical discussion that the reader without prior background in Piagetian theory can understand the issues involved. The editors have obviously tried to reach the less-advanced reader and, most especially, the educator. They have done a good job.

There are a few possible sources of confusion. For example, on page 261, "cats + noncats = 0" is not true; the intersect of cats and its complement is a null set, but the union of these two sets is the universe. '$-1 + 1 = 0$' would be a better example. But on the whole, there are very few errors or inaccuracies. The book should prove useful for a variety of readers at many levels of sophistication.

Douglass College, Rutgers University EDITH D. NEIMARK

CONTEMPORARY ISSUES IN DEVELOPMENTAL PSYCHOLOGY. Edited by NORMAN S. ENDLER, LAWRENCE R. BOULTER, and HARRY OSSER. New York: Holt, Rinehart and Winston, 1968. Pp. xvii, 682. $9.95.

What questions can a reviewer ask in evaluating a book of readings? Two are immediately obvious. He can ask whether the articles are relevant to the subject matter they are intended to discuss. Alternatively, he can ask whether the editor(s) of such a book of readings has (have) taken the care and effort to interrelate the various articles within each of the prescribed areas in such a way as to present the reader with some underlying theme or even to expose a controversy. In *Contemporary Issues in Developmental Psychology*, the editors have achieved these functions but, unfortunately, not as well as the title leads one to expect. The term 'issues' seems to imply an attempt to deal with the leading question of the day or, historically, to present various views on a single topic. It seems a bit overambitious to use the term 'issues' when, in fact, all that the editors present are areas of inquiry. Moreover, since so many of the

articles chosen for this book of readings are animal studies, the book is rather limited as a text in developmental psychology. One wonders what type of course this book could be used for. Finally, the readings are somewhat narrow. The section on perception, for example, contains one general work by Epstein, one animal experiment, and two on infancy. Surely the area of perception and development deserves a wider treatment.

The editors' handling of material, *i.e.* their comments and integrating statements, is adequate as far as it goes, but it falls short of achieving a thorough integration of the readings. After a short general introduction, the editors spend between two and four pages tying an area of inquiry together. Given the diversity of the articles within an area, these few pages can only intimate the relevance of the readings that follow.

This lack of integration and exposition, common to most books of readings, raises the far more general question of why books of readings are published and for whom they are intended. If these books are for the undergraduate, as Professor Hunt in his foreword to the present volume suggests, is there sufficient integrative material for these relatively naïve students? Without such narrative, how can they be expected to understand the complicated issues, terminology, and statistics which they encounter? Professor Hunt further suggests that such books as this are necessary to relieve the frustrating bottlenecks in the university libraries. If this is their function, would not a soft-covered and inexpensive book be more appropriate? To ask the undergraduate to spend approximately ten dollars for a single book of readings so that the library not be overburdened is too high a price, especially when it may be necessary to buy more than one book of readings for each course. Alternative methods for relieving the library need be found. With the duplicating systems now available the instructor could inform the library of the particular set of readings he requires for his course. The library (or some other publishing source) could reproduce that specific set of readings and charge the student accordingly.

Thus, if the real reason for editing and publishing such a book of readings is to aid the embattled librarian, then far cheaper means can be used. If the book is to stand only as a collection of articles to relieve the pressure at the library, why have any integrating material? If they have a greater function, then more integrating material is necessary. One might briefly consider (and reject) that such books of reading might be of use for graduate education. Imagine graduate training that denied the student the opportunity to explore an issue of a journal while looking for a particular article. We would produce a generation of students having no incidental learning. Moreover, no book of readings would have all the necessary articles for mastery of an area. Does this mean an ever-increasing library of books of readings?

It is clear that books of readings are for the most part inadequate for graduate training and that their function in undergraduate curricula needs to be questioned. Their value, then, must be more apparent to editors and publishers than to students. These comments apply to all members of the genus. The present volume is no worse than most, showing all the problems of books of readings.

MICHAEL LEWIS

Educational Testing Service
Princeton, New Jersey

MOTIVATION AND ORGANIZATIONAL CLIMATE. By GEORGE H. LITWIN and ROBERT A. STRINGER, JR. Boston: Division of Research, Graduate School of Business Administration, Harvard University. 1968. Pp. xiii, 214. $6.00.

The authors attempt to demonstrate that organizational climate directly determines motives that can be analyzed in terms of needs for achievement, power, and affiliation. These in turn involve feelings of satisfaction that result from activity of some sort (p. 9). The logic of this position makes organizational climate a set of independent variables, with feelings of satisfaction the dependent variables. Unfortunately, the authors neglect an enormous literature on job satisfaction while concentrating on the three needs. The authors write in a style that will impress readers who are naïve about the content area and about statistics. They repeatedly cite multiple references as evidence for their conclusions but frequently omit discussion of the contents of these references. This procedure is, of course, one step better than an isolated assertion or conclusion. The monograph cites an extensive literature which should be useful to the student who wishes to study some of the problems related to the three needs.

Bowling Green State University

OLIN W. SMITH
PATRICIA C. SMITH

THE DISTRIBUTION OF AUTHORITY IN FORMAL ORGANIZATIONS. By GENE W. DALTON, LOUIS B. BARNES, and ABRAHAM ZALEZNIK. Boston: Harvard University Press. 1968. Pp. xii, 229. $6.00.

Despite the somewhat imposing title, this monograph turns out to be nothing more than an evaluation of the behavioral and motivational results of an organizational change designed to "flatten" an organization. (By "flattening" an organization is meant *(a)* increasing the decision-making and planning responsibilities of those at lower levels in the organization and *(b)* decreasing the number of supervisory levels in the organization.) In essence, the design of the study reported is to describe the nature of the attempt at organizational change (an expansion of the decision-making and planning responsibilities of a group of first-line supervisors in a research and development center) and then report the results of their analysis of follow-up questionnaires designed to assess the effects of the change. Notable characteristics of the study are the utilization of comparison (not, really, control) groups for baseline comparisons of effects and separate sections indicating where the co-authors disagreed with one another as to interpretation of the findings.

On an overall basis, however, the research procedures followed were not as rigorous as could be. That is, although the findings support the utility of a "power equalization" approach to organizational performance fairly consistently, they are not of the kind to inspire great confidence, primarily because virtually complete reliance is placed on introspective descriptions of motivational change as a function of the change in the organizations. Hence, there is little or no control for the possibility that the positive results might have occurred simply as a function of the desire to balance perceptions *(e.g.* high organizational change "should" lead to high personal change, etc.) or as

a result of other such contaminating influences. Furthermore, when this weakness is combined with the reluctance of the authors to relate the results of their work to obviously relevant contemporary theoretical conceptions (*e.g.* those of McGregor, Likert, and Argyris, among others), the result is a book that is not as significant, either empirically or theoretically, as it might have been.

New York University ABRAHAM K. KORMAN

BIOLOGY OF EMOTIONS. By EDMUND JACOBSON. Springfield, Ill.: Charles C Thomas, 1967. Pp. xvii, 211. $7.50.

Since 1929 Jacobson has been arguing for the importance of neuromuscular activity in emotion. In this book Jacobson brings his thinking on this issue up to date with the attempt to present an integrated theory of emotion. The central concept of the "welfare response" is introduced and defined as the joint activity of involuntary (autonomic) and voluntary (cerebrospinal) "effort circuits" toward the accomplishment of a common adaptive goal. Emotions, in this context, are seen as welfare responses in which autonomic activity predominates. In his view, emotions can be controlled only through striated-muscle activity which is under voluntary control (Jacobson's "method of progressive relaxation") and emotion measured through the recording of the sometimes-minute muscle potentials which accompany all emotional states. In other chapters Jacobson elaborates other concepts important to this theory and reviews other theories of emotion, the neurophysiology and physiology of emotion, and animal studies (by Gellhorn), as well as some data on human subjects, in support of his general view.

Jacobson's work in this area over the past 40 years has not been generally accepted by workers in the field of emotion, nor by clinicians attempting to deal practically with emotional disorders. This may be unfortunate, but the present book will do little to change that, since the author attempts to convert rather than controvert. A balanced perspective of this field is not presented but, rather, only the evidence that generally supports his position, and at best this book serves only as a source of Jacobson's thinking about the *Biology of Emotions.*

New York University ARNOLD M. MORDKOFF

BIO-MEDICAL TELEMETRY. By R. STUART MACKAY. New York: Wiley, 1968. Pp. xi, 388. $12.50.

Electronic-circuit technology has so developed that really complicated functions can be realized in very small packages. The implications for medical, biological, and psychological research are obvious. We can now carry out a psychophysical experiment in which a whole collection of a subject's physiological variables (EEG, ECG, temperature, etc.) are measured, radioed, recorded, and processed in real time without subject constraint and even without his awareness. Dr. Mackay is a pioneer in the development of such techniques, his major interest being the "endoradiosondes" (tiny battery-powered radio transmitter-transducers ingested by or implantated in a subject

to telemeter some biological quantity) that constitute the bulk of the subject matter of this book.

Dr. Mackay is at his best when describing his own specific experiments. He presents (Chapters 1, 5–8, 13, 14, and 16) at least ten experimental case histories that include circuit design, transmitter-transducer packaging, placement on the subject, and examples of the output data; the cases are as diverse as the radio telemetry of accelerometer data from a free-roaming cockroach, the radio measurement of blood pressure with a unit implanted next to a human subject's artery, and the ultrasonic telemetry of temperature with a unit swallowed by a dolphin. Transducers are discussed for the measurement of pressure, temperature, pH, and biopotentials (ECG, EEG). Telemetry techniques considered include battery-powered radio, passive transponders, and ultrasonic waves. In Chapter 4, package protection and encapsulation is discussed. Chapters 2, 3, and 9–12 present a popularized technical treatment of transistor circuits, signal modulation, radio propagation, transmitters, and receivers; this, of course, is an immensely difficult task of exposition, and Dr. Mackay does not bring it off. He starts with basic diode action, but it is not long before terms and phrases such as "filter," "varicap diode," and "the familiar Biot-Savart law" are being thrown in with no introduction or adequate explanation. No organized discussion is offered of anatomy and physiology. No mathematical arguments or formulas are presented outside Appendixes 1 and 2 (on antennas and carrier frequencies). In fact, it is something of a puzzle to decide to just whom the book is addressed; perhaps a biologist who happens also to be an accomplished radio ham will enjoy it most.

The book has an important technical flaw in its failure to consider signal quality; the concepts of information bandwidth and signal-to-noise ratio do not even appear. What's worse, at a time when we are coming to realize the need for cooperative work by medical-psychological-engineering teams, is that the text is permeated with the atmosphere of a lone engineer off doing his thing. Dr. Mackay's interest seems to end when the gastrointestinal-pressure tracing comes off the pen recorder; we look in vain for any appraisal of the medical or biological significance of his experiments. The material is not organized well enough nor developed with a sufficiently consistent technical background to qualify as an important textbook. In effect, we have here a sort of professional autobiography of Dr. R. Stuart Mackay, with all of the advantages and disadvantages that implies.

Radio Systems Laboratory
Stanford Research Institute

WADE H. FOY

BOOKS RECEIVED

(The books listed here have not as yet been noted in our pages. Listing here does not, however, preclude their later review.)

Abelson, Robert P., and others. *Theories of Cognitive Consistency.* Chicago: Rand McNally, 1969. Pp. 920. $17.50.

Ackerknecht, Erwin H. *A Short History of Psychiatry.* 2nd rev. ed., transl. by Sula Wolff. New York: Hafner, 1968. Pp. 109. $4.25.

Anderson, Bert D. *Introduction to College.* New York: Holt, Rinehart and Winston, 1969. Pp. 241. $3.95.

Arnold, Magda B. *The Nature of Emotion.* Baltimore: Penguin Books, 1968. Pp. 398. $1.95.

Baier, Kurt, and Rescher, Nicholas. *Values and the Future.* New York: The Free Press, 1969. Pp. 527. $14.95.

Bennis, Warren G., Benne, Kenneth D., and Chin, Robert. *The Planning of Change.* 2nd ed. New York: Holt, Rinehart and Winston, 1969. Pp. 627. $8.95.

Berman, Louis A. *Jews and Intermarriage: A Study in Personality and Culture.* Cranbury, N.J.: Thomas Yoseloff, 1968. Pp. 707. $12.00.

Borgatta, Edgar F., and Evans, Robert R. *Smoking, Health, and Behavior.* Chicago: Aldine, 1969. Pp. 288. $10.75.

Borgatta, Edgar F., and Lambert, William W. *Handbook of Personality Theory and Research.* New York: Rand McNally, 1969. Pp. 1,216. $20.00.

Bormann, Ernest G. *Discussion and Group Methods: Theory and Practice.* New York: Harper & Row, 1968. Pp. 340. $8.95.

Bower, Eli M. *Early Identification of Emotionally Handicapped Children in School.* 2nd ed. Springfield, Ill.: Charles C Thomas, 1969. Pp. 261. $9.75.

Bowman, Jan C. *The Figure-Ground Phenomenon in Experimental and Phenomenological Psychology.* Stockholm: University of Stockholm, 1968. Pp. 269.

Buss, Arnold H., and Buss, Edith H., eds. *Theories of Schizophrenia.* New York: Atherton Press, 1969. Pp. 175. Cloth, $6.95; paperback, $2.95.

Butcher, James N. *MMPI: Research Developments and Clinical Applications.* New York: McGraw-Hill, 1969 Pp. 402. $9.95.

Canestrelli, L., Chavin, R., Colle, J., Friasse, P., Klineberg, O., Musetti, C., and Zazzo, R. *Le comportement.* Paris: Presses Universitaires de France, 1968. Pp. 171.

Carroll, Herbert. *Mental Hygiene.* 5th ed. Prentice-Hall, 1969. Pp. 436. $8.95.

Chomsky, Noam. *Aspects of the Theory of Syntax.* Cambridge: The M.I.T. Press, 1969. Pp x, 251. Cloth, $7.50; paperback, $2.95.

Cratty, Bryant J. *Perceptual-Motor Behavior and Educational Processes.* Springfield, Ill.: Charles C Thomas, 1969. Pp. 265. $9.50.

Dolezel, Lubomir, and Bailey, Richard W. *Statistics and Style.* New York: American Elsevier, 1969. Pp. 245. $9.75.

Dustin, David S. *How Psychologists Do Research: The Example of Anxiety.* Englewood Cliffs, N.J.: Prentice-Hall, 1969. Pp. 108. Paperback, $2.50.

Edwards, Allen L. *Statistical Analysis.* 3rd ed. New York: Holt, Rinehart and Winston, 1969. Pp. 244. $7.95.

Eklund, Harold. *Multiple Choice and Retention.* Stockholm: Larome Delsforlagen, 1968. Pp. 223.

Elind, D., and Flavell, J., eds. *Studies in Cognitive Development.* New York: Oxford University Press, 1969. Pp. 503. Cloth, $10.00; paperback, $4.50.

Escalona, Sibylle K. *The Roots of Individuality.* Chicago: Aldine, 1969. Pp. 547. $14.75.

Etzioni, Amitai, ed. *The Semi-Professions and Their Organization.* New York: The Free Press, 1969. Pp. xix, 328. $10.95.

Falck, Frank J. *Stuttering.* Springfield, Ill.: Charles C Thomas, 1969. Pp. 160.

Feuer, Lewis S. *The Conflict of Generations.* New York: Basic Books, 1969. Pp. 543. $12.50.

Foss, Brian M., ed. *Determinants of Infant Behavior.* Vol. 1. *Proceedings of the Tavistock Seminar on Mother-Infant Interaction.* New York: Barnes & Noble, 1969. Pp. 304. $8.00.

Francis-Williams, Jessie. *Rorschach with Children.* New York: Pergamon Press, 1969. Pp. 168. Cloth, $7.00; paperback, $4.45.

Furth, Hans G. *Piaget and Knowledge.* Englewood Cliffs, N.J.: Prentice-Hall, 1969. Pp. 270. $7.95.

Galanter, Eugene. *Textbook of Elementary Psychology.* San Francisco: Holden-Day, 1966. Pp. 419. $8.50.

Garaudy, Roger. *From Anathema to Dialogue.* New York: Random House, 1968. Pp. 124. $1.45.

Gathercole, C. F. *Assessment in Clinical Psychology.* Baltimore: Penguin Books, 1968. Pp. 174. $1.45.

Gelbaum, Bernard R., and March, James G. *Mathematics for the Social and Behavioral Sciences: Probability, Calculus, and Statistics.* Philadelphia: W. B. Saunders, 1969. Pp. 337. $8.75.

Ginott, Haim G. *Between Parent and Teenager.* New York: Macmillan, 1969. Pp. 256. $5.95.

Glass, David C., ed. *Environmental Influences.* New York: Rockefeller University Press and the Russell Sage Foundation, 1968. Pp. 304. $7.50.

Goldman, George D., and Milman, Donald S. *Modern Woman: Her Psychology and Sexuality.* Springfield, Ill.: Charles C Thomas, 1969. Pp. 271. $10.50.

Gould, Laurence J., and Martin, William G. *Think about It.* New York: Prentice-Hall, 1969. Pp. 55. $3.25.

Grey, Alan L., ed. *Class and Personality in Society.* New York: Atherton Press, 1969. Pp. 190. Cloth, $6.95; paperback, $2.95.

Grollman, Earl A. *Explaining Divorce to Children.* Boston: Beacon Press, 1969. Pp. 257. $5.95.

Guerney, Bernard G., Jr. *Psychotherapeutic Agents: New Roles for Nonprofessionals, Parents, and Teachers.* New York: Holt, Rinehart and Winston, 1969. Pp. 595. $5.95.

Guillaumin, Jean. *La genese du souvenir.* Paris: Presses Universitaires de France, 1968. Pp. 307. $5.00.

Harris, D. H., and Chaney, F. B. *Human Factors in Quality Assurance.* New York: Wiley, 1969. Pp. 234. $10.95.

Hart, Benjamin L. *Experimental Neuropsychology: A Laboratory Manual.* San Francisco: W. H. Freeman, 1969. Pp. 102. $4.75.

Harvard Educational Review. *Equal Educational Opportunity.* Cambridge: Harvard University Press, 1969. Pp. 274. $5.95.

Hellmuth, Jerome, ed. *The Disadvantaged Child.* Vol. 2. Seattle: Special Child Publications, 1969. Pp. 621. $12.50.

Hill, John P., ed. *Minnesota Symposia on Child Psychology.* Vol. 2. Minneapolis: University of Minnesota Press, 1969. Pp. 172. $5.00.

Holt, John. *How Children Learn.* New York: Pitman, 1969. Pp. 192. $2.25.

Israel, Lucien. *Le medecin pace au malade.* Brussels: Charles Dessart, 1968. Pp. 338.

Joch, Hans H. *Violent Men.* Chicago: Aldine, 1969. Pp. 268. $7.50.

Johnson, Ronald C., and Medinnus, Gene R., eds. *Child Psychology: Behavior and Development.* 2nd ed. New York: Wiley, 1969. Pp. xi, 677. $8.95.

Jones, Maxwell. *Social Psychiatry in Practice.* Baltimore: Penguin Books, 1969. Pp. 190. $1.25.

Kadushin, Charles. *Why People Go to Psychiatrists.* New York: Atherton Press, 1962. Pp. 373. $8.95.

Kellogg, Rhoda. *Analyzing Children's Art.* Palo Alto, Calif.: National Press, 1969. Pp. 308. $8.95.

Kiesler, C. A., Collins, B. E., and Miller, N. *Attitude Change.* New York: Wiley, 1969. Pp. 386. $8.95.

Kitano, Harry H. L. *Japanese Americans: The Evolution of a Subculture.* Englewood Cliffs, N.J.: Prentice-Hall, 1969. Pp. 186. $5.95.

Krech, D., and others. *Elements of Psychology.* 2nd ed. New York: Random House, 1969. Pp. 864, index.

Lana, Robert E. *Assumptions of Social Psychology.* New York: Appleton-Century-Crofts, 1969. Pp. 179. Cloth, $5.00; paperback, $2.75.

Lindinger, H. C. *Familienumwelt und Prognose der Schizophrenie: Die Bedeutung verschiedener Familiensituationen fur den Verlauf.* Berne: Hans Huber, 1968. Pp. 246. DM 24.

Louch, A. R. *Explanation and Human Action.* Berkeley and Los Angeles: University of California Press, 1969. Pp. 243. $2.65.

Lunzer, E. A., and Morris, J. F., eds. *Contexts of Education. (Development in Learning,* Vol. 3.) New York: American Elsevier, 1969. $10.50.

McGhie, Andrew. *Pathology of Attention.* Baltimore: Penguin Books, 1969. Pp. 192. $1.65.

McLaughlin, Barry. *Studies in Social Movements.* New York: The Free Press, 1969. Pp. 497. $8.95.

Maher, Brendan, ed. *Clinical Psychology and Personality: The Collected Papers of George Kelly.* New York: Wiley, 1969. Pp. 361. $8.95.

Menyk, Paula. *Sentences Children Use.* Cambridge: The M.I.T. Press, 1969. Pp. 165. $7.50.

Meynaud, Jean. *Technocracy.* Transl. by Paul Barnes. New York: The Free Press, 1969. Pp. 315. $12.95.

Millar, Susanna. *The Psychology of Play.* Baltimore: Penguin Books, 1968. Pp. 288. $1.45.

Minsky, Marvin, and Papert, Seymour. *Perceptrons.* Cambridge: The M.I.T. Press, 1969. Pp. 258. Cloth, $12.00; paperback, $4.95.

Minsky, Marvin, ed. *Semantic Information Processing.* Cambridge: The M.I.T. Press, 1968. Pp. 438. $15.00.

Moskowitz, Merle J., and Orgel, Arthur R. *General Psychology.* Boston: Houghton Mifflin, 1969. Pp. 450. $6.95.

Munn, Norman L., Fernald, L. Dodge, Jr., and Fernald, Peter S. *Basic Psychology.* 2nd ed. Boston: Houghton Mifflin, 1969. Pp. 557. $6.25.

Munn, Norman L., Fernald, L. Dodge, Jr., and Fernald, Peter S. *Introduction to Psychology.* 2nd ed. Boston: Houghton Mifflin, 1969. Pp. 752. $9.50.

Mussen, Paul Henry, Conger, John Haneway, and Kagan, Jerome. *Child Development and Personality.* 3rd ed. New York: Harper & Row, 1969. Pp. 795. $9.95.

O'Neil, W. M. *The Beginnings of Modern Psychology.* Baltimore: Penguin Books, 1968. Pp. 157. $1.25.

Ost, Allison, and Restle, Vance. *A Laboratory Introduction to Psychology.* New York: Academic Press, 1969. Pp. 214. $4.25.

Phillips, John L., Jr. *The Origins of Intellect: Piaget's Theory.* San Francisco: W. H. Freeman, 1969. Pp. xviii, 149. Paperback, $2.75.

Reissman, Frank. *Blueprint for the Disadvantaged.* New York: Anti-Defamation League, 1968. Pp. 43. $0.75.

Rieselbach, Leroy N., and Balch, George I. *Psychology and Politics: An Introductory Reader.* New York: Holt, Rinehart and Winston, 1969. Pp. 305. Paperback, $4.50.

Robaye, Francine. *L'enfant au cerveau blessé.* Brussels: Charles Dessart, 1969. Pp. 222.

Rodin, Ernst A. *The Prognosis of Patients with Epilepsy.* Springfield, Ill.: Charles C Thomas, 1969. Pp. 455. $19.50.

Rogers, Dorothy. *Issues in Adolescent Psychology.* New York: Appleton-Century-Crofts, 1969. Pp. 614. Paperback, $4.95.

Rosenhan, David, and London, Perry, eds. *Theory and Research in Abnormal Psychology.* New York: Holt, Rinehart and Winston, 1969. Pp. 552. Paperback, $6.95.

Rosenthal, David, and Kety, Seymour S. *The Transmission of Schizophrenia.* London: Pergamon Press, 1969. Pp. 433. $15.00.

Rubin, Theodore Isaac. *The Angry Book.* New York: Macmillan, 1969. Pp. 179. $4.95.

Rycroft, Charles, Gorer, Geoffrey, Storr, Anthony, Wren-Lewis, John, and Lomas, Peter. *Psychoanalysis Observed.* Baltimore: Penguin Books, 1968. Pp. 160. $1.65.

Sahakian, William S., ed. *Psychotherapy and Counseling Studies in Technique.* New York: Rand McNally, 1969. Pp. 592. $7.95.

Sarason, Seymour B., and Doris, John. *Psychological Problems in Mental Problems.* 3rd ed. New York: Harper & Row, 1968. Pp. 678. $9.95.

Savage, R. Douglass. *Psychometric Assessment of the Individual Child.* Baltimore: Penguin Books, 1968. Pp. 117. $1.25.

Schlesinger, I. M. *Sentence Structure and the Reading Process.* Paris: Mouton, 1968. Pp. 171.

Schultz, Duane P. *A History of Modern Psychology.* New York: Academic Press, 1969. Pp. 346. $7.50.

Semidt, J. E. *Paramedical Dictionary: A Practical Dictionary for the Semi-Medical and Ancillary Medical Professions.* Springfield, Ill.: Charles C Thomas, 1969. Pp. 423. $8.75.

Searle, John R. *Speech Acts: An Essay in the Philosophy of Language.* New York: Cambridge University Press, 1969. Pp. 203. $7.50.

Sherwin, Robert Veit. *Compatible Divorce.* New York: Crown Publishers, 1969. Pp. 308. $5.95.

Siegel, Laurence. *Industrial Psychology.* Homewood, Ill.: Richard D. Irwin, 1969. Pp. 594. $9.50.

Simirenko, Alex, ed. *Social Thought in the Soviet Union.* Chicago: Quadrangle Books, 1969. Pp. 439. $14.95.

Smith, W. Porter. *General Psychology.* New York: McGraw-Hill, 1969. Pp. 381. $7.95.

Snider, James G., and Osgood, Charles E. *Semantic Differential Technique.* Chicago: Aldine, 1969. Pp. 681. $12.50.

Stein, Calvert. *Practical Psychotherapy in Nonpsychiatric Specialties.* Springfield, Ill.: Charles C Thomas, 1969. Pp. 300. $10.75.

Sullivan, Vincent F. *How to Stop Problem Drinking.* New York: Frederick Fell, 1969. Pp. 228. $5.95.

Talland, George A. *Disorders of Memory and Learning.* Baltimore: Penguin Books, 1969. Pp. 176. $1.45.

Tavolga, William N. *Principles of Animal Behavior.* New York: Harper & Row, 1969. Pp. 143. Paperback, $4.00.

Vincent, Clark E. *Unmarried Mothers.* New York: The Free Press, 1969. Pp. 308. $2.45.

von Bertalanffy, Ludwig. *General System Theory.* New York: George Braziller, 1969. Pp. 289. Cloth, $8.95; paperback, $3.95.

Van Hattum, Rolland J., ed. *Clinical Speech in the Schools: Organization and Management.* Springfield, Ill.: Charles C Thomas, 1968. Pp. 381. $12.50.

Wolfle, Dael, ed. *The Discovery of Talent.* Cambridge: Harvard University Press, 1969. Pp. 316. $9.50.

Wrage, Karl Horst. *Children: Choice or Chance.* Philadelphia: Fortress Press, 1969. Pp. 119. $2.95.

Wright, George N., ed. *Wisconsin Studies in Vocational Rehabilitation.* Madison: University of Wisconsin Press, 1968.

Zaetz, Jay L. *Occupational Activities Training Manual for Severely Retarded Adults.* Springfield, Ill.: Charles C Thomas, 1969. Pp. 107. $7.25.

CORRECTION

The preliminary footnote to "The Absolute and Relative Size Cues to Distance" (Vol. 82, No. 2, p. 228) should read:

* Received for publication August 14, 1967. The study was supported by Grant NGR 05-010-010 from the National Aeronautics and Space Administration. The author wishes to thank Steven Foote, David Reese, and David Schultz for their help in collecting the data.

The preliminary footnote to "Equidistance Effects in Visual Fields" (Vol. 82, No. 3, p. 342) should read:

* Received for publication January 8, 1968. The study was supported by Grant NGR 05-010-010 from the National Aeronautics and Space Administration. The author wishes to thank Steven Foote, Michael Koslow, Robert Newton, David Reese, Robert Simpson, and Thomas Zallen for their help in constructing the apparatus and collecting the data, and Meredith Puterbaugh for her assistance in the data analysis.

INDEX

Rand B. Evans, Wright State University

AUTHORS

(The names of authors of original articles are printed in CAPITALS and SMALL CAPITALS; of authors of books reviewed, in roman; and of reviewers, in *italics*.)

Aboud, J., Jr. 221
Adkins, M. 122
Andrews, O. E. . . . 523
Anstis, S. M. 350, 474
Antonitis, J. J. . . . 389

Baer, P. E. 168
Bagley, S. K. 307
Barnes, L. W. 554
Beck, J. 359
Benton, R. G. 133
Berger, S. M. 333
Bevan, W. 415
Bitterman, M. E. 126
Boulter, L. R. 552
Bowman, R. W., Jr. 23
Brehmer, B. 490
Burnham, C. A. . . . 86

Campione, J. C. . . . 142
Campione, J. C. . . 212
Ceraso, J. 548
Cherry, Colin 551
Chubb, N. C. 307
Clausen, J. 139
Cofer, C. N. 549
Cohen, J. 550
Collyer, S. C. 415
Cronnell, B. 248

Dalton, G. W. 554
Daves, W. F. 122
Day, R. H. 191

Ellis, N. R. 139
Ellsbury, S. W. . . 333
Endler, N. W. 552
Eriksen, C. W. . . . 320

Fellows, B. J. 142
Filion, R. D. L. . . 226
Fishbein, H. D. . . 523
Fishbein, M. 457
Fowler, S. C. 266
Foy, W. H. 555

Freund, J. S. 198
Fuhrer, M. J. 168

Glass, T. A. 241
Glucksberg, S. . . . 111
Gogel, W. C. . . 228, 342
Granberg, D. 221
Greenberg, B. 254

Halper, F. 425
Hamilton, D. 272
Harcum, E. R. 504
Hatch, G. 457
Hooper, F. H. 552
Hurst, P. M. 307

Jacobson, E. 555
Johnson-Laird, P. N. 367
Jung, J. 548

Kaplan, M. F. 157
Kappauf, W. E. 40, 547
Karkau, V. T. 203
Kennedy, R. H. . . 272
Kohfeld, D. L. . . . 104
Kohler, Y. 141
Korman, A. K. 554
Kovar, L. C. 143

Landy, E. 457
Langer, J. 181
Lasker, A. 1
Lawrence, D. H. . . . 549
Levine, M. 117
Levinthal, C. F. . . 523
Lewis, M. 552
Litwin, G. H. 554
Lott, D. 248
Lovejoy, E. 549
Lovelace, E. A. . . . 254
Lowry, R. 391
Lytle, L. D. 111

MacKay, D. G. 23

Mackay, R. S. 555
Mandler, G. 531
Mancy, P., Jr. 133
Mayhew, J. W. . . . 474
Michon, J. A. 550
Mordkoff, A. M. . . . 555
Morgan, M. J. 380
Morgan, R. 117
Morley, T. 474
Mundl, W. J. 406

Neiberg, A. 117
Neimark, E. D. 552
Neisser, U. 540
Newman, E. B. . . . 400
Notterman, J. M. 266

Olczak, P. V. 157
Ono, H. 86
Osser, H. 552

Posner, M. 551
Powledge, F. 143

Rabinowitz, F. M. 142
Radlow, R. 307
Rappoport, D. A. . . 133
Robertson, B. 133
Robinson, J. S. . . . 513
Roby, T. B. 374
Rock, I. 1, 425
Rock, I. 141
Rosenberg, B. G. . 181
Royer, J. M. 96
Rump, E. E. 291

Schiff, W. 73
Schurman, D. L. . . 320
Siegel, F. H. 552
Simon, J. 1
Sjöberg, L. 441
Skinner, B. F. 549
Smith, F. 241
Smith, O. W. 554
Smith, P. C. 554

STEGER, J. A. 56
STEWART, J. 181
Stringer, R. A., Jr. 551
SUMMERS, D. A. . . . 235
SUMMERS, R. C. . . . 203
SUMMERS, S. A. . . . 203

TAGART, J. 367
TARPY, R. M. 111
TAYLOR, R. L. 259
THAYER, S. 73

UNDERWOOD, B. J. . . 198

WADE, N. J. . . . 191, 384
WHITE, J. F. 124
WILSON, J. A. 350

Zaleznik, A. 554

SUBJECTS

(The references in *italic* figures are to reviews.)

Adaptation: perceptual, 141 f.; proprioceptive, visual-orientation judgments and, 195 ff.; shifts of cue(s) and rules in multiple-probability tasks, 235–240
Adaptation-level effects, simple reaction time and, 104–109
Aftereffect, visual tilt: lateral body tilt and, 191–197
Animals: monkey, cutaneous feedback and motor performance in, 270 f.; rat, degree of exposure and aversiveness of nonreward cues in, 279–282; conditioning of, apparatus for, 133–136; small, apparatus for reward testing with, 406–410
Apparatus: conditioning, for aquatic animals, 133–136; conditioning, testing of olfactory stimuli in, 133–136; food dish, for reward testing with animals, 406–410; hurdle, for training aquatic animals, 133–136; randomizer, for forced-choice visual stimuli, 410–413; telemetry, biomedical, *555 f.*
Apprehension, visual: of numerousness, context effects in, 221–227
Assimilation, lightness: differential stimulation and, 56–72
Association, word: original thinking and, 157–167
Authority in formal organizations, *554 f.*
Autokinesis, Myers' phenomenon of, *124 ff.*
Aversiveness of nonreward cues, degree of exposure and, 279–282

Biography: C. Lloyd Morgan, 126–133
Brain stimulation, electrical; grid-delivered shock and, 136 ff.

Cartesian psychology, 393
Children, thinking in, *552*
Cognition: concept identification and hypothesis-testing behavior, 523–530; concept shifts in retarded children, 212–220; concept shifts, theory of, 219 f.; cue-criterion relations in cognition, 490–503; effect of, on differential GSR conditioning, 168–180; learning task, as measure of, 493–503; task characteristics and, 490–503; understanding implications and, 367–373. *See also* Thinking
Communication: human, *551 f.*; sensory, *551 f.*
Concept identification: organization of hypothesis-testing behavior and, 523–530; stimulus dimensions and, 523–530
Concept shifts: in retarded children, 212–220; theory of, 219 f.
Conditioning: apparatus for, in aquatic animals, 133–136; GSR, cognitive process in, 168–180; masking task and, 168–180; probability-learning task and, 168–180
Congruency in symbolic learning, 182–190
Constancy principle, Helmholtz', 397 ff.
Context: effect of, in judgment of visual numerousness, 221–227; invariant, performance in paired-associates learning with, 423 f.; thematic relevance of, in paired-associates learning, 415–424; verbal, in paired-associates learning, 415–424
Criteria, function relationships between cues and, 203–211
Cues: association of, with nonreward, 279–282; functional relationships between criteria and, 202–211; nonreward, aversiveness of, 279–282; optical contact as, in perception of size and distance, 342–349;

size judgments, distance judgments and, 228–234
Cue-criterion relations in judgment, 490–503
Cue selection: length of study interval and, in paired-associates learning, 254–258; retroactive inhibition and, in interpolated learning, 278 f.
Cue shift, adaptation to: in multiple-probability tasks, 235–240
Cutaneous feedback, changes in: motor-task performance and, 266–269, 271; in monkey, 270 f.

D-amphetamine, effects of: on acquisition of new material in paired-associates learning, 309–313, 316–319; on delayed recall in paired-associates learning, 314–319; on persistence of paired-associates learning, 313 f., 316–319; on recall, 309–319
Descartes, René: influence of Galilean physics on ideas of, 394 f.
Development, child: in school environment, *143 f.*; in discrimination learning, *142 f.*; learning and, *143 f.*
Developmental psychology, *552 f.*
Dimensionality of judgments of similarity, 441–456
Discrimination, process of: development and, *142 f.*
Distance, judgment of: cues to size and, 228–234; judgment of size, invariance hypothesis and, 291–306; visual-field changes, effect of on, 342–349
Distortion, perceptual, *141 f.*

Eckman's law, statement of, 305 f.
Emotion: biology of, *555*; judgment of, in facial expressions, 441–456
Equidistance tendency in visual fields, 333–349
Estimates. *See* Judgment
Event cycles, prediction of, 374–379
Expression, facial: similarity judgments of, 444–456
Extinction, latent: frustration hypothesis of, 280; in rats, 279–282
Facial expression, judgment of social interactions by, 73–85
Feedback: changes in cutaneous, motor-task performance and, 266–269, 271; changes in cutaneous, motor performance in monkey and, 270 f.; changes in visual, motor-task performance and, 266 f., 269 ff.

Figures: ambiguous, of concavity, 389 ff.; ambiguous, of convexity, 389 ff.; recall of, differences of size and shape in, 7–22; recognition of, differences of size and shape in, 7–22
Freudian theory: influence of Newton on, 398 f.; influence of Helmholtz on, 398 f.

Galileo, influence of: on Hobbes, 394 f.; on Descartes, 394 f.; on psychological thought, 391–400
Galvanic Skin Response (GSR): conditioning of, cognitive processes in, 168–180; conditioning of, masking task and, 168–180; conditioning of, probability-learning task and, 168–180; differential, accuracy of verbalization and, 168–180
Generalization, stimulus: as recognition process, 1–22; theory of, 1–4, 19–22; unawareness of stimuli differences in, 3; vs. associative theory, 4 f., 7–22
Gestalt psychology, dependence on physicalism in, 400 f.
GSR. *See* Galvanic Skin Response

Helmholtz, Hermann: constancy principle of, 397 ff.; contributions to psychology by, 402; influence of, on Freudian theory, 398 f.
Hobbes, Thomas: influence of Galilean physics on ideas of, 394 f.; psychological systems of, compared to Descartes', 393 f.
Hypothesis testing, concept identification and, 523–530

Illusion: autokinetic, Myers', 124 ff.; Müller-Lyer, Fisher's explanation of, questioned, 380 f.; Müller-Lyer, Gregory's explanation of, questioned, 381 f.; Müller-Lyer modified, estimates of length and subjective geometry in, 380–384, 383 f.; photographic, of convexity and concavity, 389 ff.; photographic, reversible, 389 ff.; trapezoidal window, effect of action on, 86–95; trapezoidal window, effect of directed attention on, 89 f., 93 ff.; trapezoidal window, variables influencing, 86–95
Incentive, effect of: on short-term recall, 111–116
Inference: Brunswik's lens model of, 492; clinical, information utilization in, 498–503; cue-criterion re-

lations in. 490; information processing in, 500–503
Information: processing of, in inference, 500–503; stored vs. sensory, in recognition, 259–265
Intrusion, semantic: in bilingual transfer, 33 f.
Invariance hypothesis: Cook's experiment on, 291–295; Cook's experiment replicated, 291–306

Judgment: cue-criterion relations in, 490–503; of difference, stored vs. sensory information in, 259–265; of direction of gaze in another person, 474–489; of distance, size cues and, 228–234; of emotions in facial expressions, 444–456; of social interaction, motion patterns and, 73–85; of visual numerousness, contextual effects in, 221–227; orientational-visual, reference for, 191–197; process of, 491–503; process of, characteristics of task and, 491–503; relational, 203–211; relational, between objects, 73–85; similarity, dimensionality in, 441–456; similarity, in facial expressions, 444–456; similarity, methods of determining and comparisons thereof, 441 f., 448, 454 ff.; similarity, scaling of, 441–456; size-distance, changes in visual field and, 342–349; size-distance, invariance hypothesis and, 291–306; social interaction, facial expressions and, 73–85; visual-orientational, ocular countertorsion and, 195 ff.; visual-orientational, proprioceptive adaptation and, 195 ff.; visual-orientational, references for, 191–197

Leadership, effectiveness of: Fiedler's model of, 457 f.; Fiedler's model, assumptions underlying, 458–473; group-task situations and, 466 f., 473; interactions among variables in, 469–473; leader-member affective relationships and, 467, 473; task structure and, 467 f., 473
Learning: discrimination, models of, *549*; free-recall, effect of list organization on strategies in, 531–539; incidental, effect of verbal reinforcement on, 333–341; incidental vs. intentional in orienting task, 333–341; instrumental, in chicks, 126 f.; instrumental, C. Lloyd Morgan's theory of, 126–133; intentional vs. incidental in orienting task, 333–341; interpolated, cue selection in and retroactive inhibition, 278 f.; method, effectiveness of interpolated taks, 111–116; nonverbal, 181–190; nonverbal, congruency in, 182–190; nonverbal, gestures as, 185–190; nonverbal, sounds in, 185–190; nonverbal, verbal referents and, 182 ff., 188 ff.; paired-associates, invariant context and, 415 f.; paired-associates, thematic relevance of invariant context and, 415–424; paired-associates, cue selection and, 276–279; paired-associates, cue selection, length of study interval and, 254–258; paired-associates, cue selection, interpolated learning, retroactive inhibition and, 278 f.; paired-associates, *d*-amphetamine and, 307–319; paired-associates, *d*-amphetamine, acquisition of new material and, 309–313, 316–319; paired-associates, delayed recall and effects of, 314–319; paired-associates, effect of delayed recall on persistence in, 313 f., 316–319; paired-associates, meaningful stimuli and, 415–424; paired-associates, practice and, 276–279; paired-associates, recall as prerequisite for recognition in, 96–103; paired-associates, recognition as prerequisite for recall in, 1–22, 96–103; paired-associates, retroactive inhibition and, 276–279; paired-associates, verbal context and, 415–424; probability, GSR conditioning and, 168–180; multiple-probability, adaptation to cue shifts in, 235–240; multiple-probability, adaptation to rule shifts in, 235–240; relational, semantic, 181–190; symbolic, 181–190; symbolic, congruency in, 182–190; symbolic, of gestures, 185–190; symbolic, of sounds, 185–190; symbolic, verbal referents and, 182 ff.; verbal, 548 f.; verbal, discriminative, number of incorrect terms and, 198 ff. *See also* Memory; Recall; Recognition
Learning theory, discrimination, 142 f.
Length, estimates of: in modified Müller-Lyer illusion, 380–384
Letters: identification of, distinctive-feature theory of, 248–252 f.; recognition of, 259–265

Lightness assimilation, differential stimulation and, 56–72
Lightness contrast: differential stimuli and, 56–72; Helson's theory of, 56 f., 69–72
Lightness, perceived: apparent depth and, 363–366; figural orientation and, 359–366
Linguistics, meaning in bilingual transfer, 23–39
Lists, length of: digit retrieval and, 272–275
Locke, John: influences of Newton on, 392

Masking: GSR conditioning in paradigm of, 168–180; visual, failure to replicate disinhibition of, 320–332
Meaning: in bilingual transfer, 23–29; production of, in sentences, 23–39; symbolic, 181–190; transfer of, in sentences by bilinguals, 23–39
Meetings: Society of Experimental Psychologists, 547; Southern Society for Philosophy and Psychology, 413 f.
Memory: short-term, covert rehearsal and, 111–116; short-term, incentive and, 111–116; verbal, rhythmic grouping structure and, 540–546; verbal, role of rhythm in serial intrusion in, 540–546
Memory span, verbal: rhythmic grouping in, 540–546; role of rhythm in serial intrusions in, 540–546
Memory trace in recognition, 1–22
Model: Brunswik's lens, of inference, 492 f.; Fiedler's, for leadership effectiveness, 471 f.; physicalist, in psychology, 400–406
Morgan, C. Lloyd: influence of, in instrumental learning, 126–133
Motivation: organizational climate and, 554; covert rehearsal and, in short-term memory, 111–116
Motor task, performance in: cutaneous feedback and, 266–269, 271; visual feedback and, 266 f., 269 ff.; weight of limb and, 270 f.

Newton, Isaac, influence of: on John Locke, 392; on nineteenth-century psychology; 400–406; on psychological thought, 391–400
Normalization, visual: reference for, 191–197; retinal vs. gravitational postural adaptation and, 194–197
Numerousness, visual: contextual effects in judgment of, 221–227

Optical contact, as cue in perception of size and distance, 333–349
Orientational-visual judgments, ocular countertorsion and, 195 ff.
Orientation, visual: monocular vs. binocular observation and, during head tilt, 384–388
Originality: study of, 157–167; theories of, 157
Perception: of direction of gaze, 474–489; of extensity, phenomenal vs. retinal location of points in, 425–440; of form, phenomenal vs. retinal location of points in, 425–440; of form, without retinal image, 425–440; of length, in modified Müller-Lyer illusion, 380–384; pattern, novel vs. familiar stimuli in, 513–522; pattern, primacy effect in, with binocular tachistoscopic viewing across fixation, 504–512; of size and distance, invariance hypothesis and, 291–306; space, 141 f.; of time, in health and disease, *550 f.*; visual delay, luminance level and, 350–358; visual field, equidistance effects in, 342–349; visual invariance hypothesis, replication of Cook's experiment on, 291–306; visual lightness, apparent depth and, 363–366; visual lightness, assimilation effects in, 56–72; visual lightness, contrast effect, 56–72; visual lightness, Helson's theory of, 56 f., 69–72; visual lightness, figural orientation and, 359–366; visual masking, 320–332; visual masking, cues in, 327–332; visual verticality, monocular vs. binocular observation and, during head tilt, 384–388
Perceptual adaptation, *141 f.*
Perceptual distortion, *141 f.*
Performance, motor-task: cutaneous feedback and, 266–269, 271; kinesthetic feedback and, 270 f.; visual feedback and, 266 f., 269 ff.
Practice: speech rate in bilinguals and, 23 ff., 28–32; bilingual transfer and, with delayed auditory feedback, 32–35; meaning production and, in sentences, 23–39
Preference: delay, pleasantness of waiting period and, in children, 241–247; delay, uncertainty of score and, in children, 241–247; delay, vividness of score presentation and, in children, 241–247; dimensional, in concept shifts of retarded children, 212–220

Problem solving, concept shifts in: with retarded children, 212–220; theory of, 219 f.
Proprioception, adaptation in: visual-orientation judgments and, 195 ff.
Psychology, physicalist models of, 400
Psychophysics: apparatus for forced-choice procedure, stimulus randomization in, 410–413; up-and-down method of, test of Dixon and Mood statistics for, 40–55

Reaction time: as measure of recognition, 260–265; digit retrieval and, 272–275; original thinking and, 158–167; simple, adaptation-level phenomenon in, 104–109; simple, ready-signal intensity and, 104–109; simple, sequential intensity effects on, 104–109
Recall: associative, stimulus recognition and, 96–103; figural, size and shape differences in, 7–22; free, saturation effects in, 122 ff.; free, stimulus variation and, 122 ff.; as prerequisite for recognition, in paired-associates task, 96–103; recognition and, interactions between, 259–265; short-term, covert rehearsal in, 111–116; short-term, effectiveness of interpolated-task procedure on, 111–116; short-term, effectiveness of incentives on, 111–116
Recognition: figural, size and shape differences in, 7–22; interaction between stored and sensory information in, 259–265; of letters, 259–265; process of, 1 ff.; as requirement for recall in paired-associates learning, 1–22, 96–103; stimulus generalization as process of, 1–22; visual, interactions between stored and sensory information in, 259–265
Rehearsal, covert: in short-term memory, 111–116; incentives and, 111–116; interpolated-tasks procedure and, 111–116
Reinforcement: delayed, preference for task and, in children, 241–247; expressive verbal, effect on incidental learning, 333–341; immediate, preference for task and, in children, 241–247; to cutaneous pressure bands, in squirrel monkey, 270 f.
Relation, judgments of, 202–211
Remote Associates Test as measure of originality, 157 ff.
Retardation, research in, *139 ff.*
Retrieval, digit: serial position in list and, 272–275; length of list and, 272–275
Retroactive inhibition: cue selection in interpolated learning and, 278 f.; in paired-associates learning, 276–279
Rule shift, adaptation to: in multiple-probability tasks, 235–240

Scaling, similarity: category rating and, 415 ff., 449–456; magnitude estimation and, 415–449, 453–456; paired comparisons in, 415–449, 455 f.; unidimensional, differing response modes and, 451–456
Search, visual: type case and efficiency in, 248–253; type size and efficiency in, 248–253
Semantic errors in bilingual transfer, 33 f.
Semantic relationships, learning of, 181–190
Sequences of nonsense syllables, prediction of, 374–379
Shock, grid-delivered: with electrical brain stimulation, 136 ff.
Size: absolute vs. relative, as cues for distance judgments, 228–234; of type, word identification and, 248–253
Size-distance judgments: changes in visual field and, 342–349; invariance hypothesis and, 291–306; optical contact as a cue in, 342–349
Social interaction, judgment of: facial expressions and, 73–85
Speech, errors of: delayed auditory feedback and, in bilingual transfer, 32–35
Speech production, bilingual: role of meaning in, 23–39
Speech rate: familiarity with meaning of sentences and, in bilinguals, 25 f., 28–32; practice and, in bilinguals, 23 ff., 28–32
Spoonerisms in bilingual transfer, 34 f.
Statistics, Dixon and Mood's, for up-and-down psychophysical method, 40–55
Study interval, length of: cue selection and, in paired-associates learning, 254–258

Teaching: methods of, *142 f.;* practice of, *549 f.*
Telemetry, biomedical, *555 f.*
Thematic Apperception Test (TAT),

originality and behavior on, *160–167*
Thinking: in children, *552*; original, 157–167; theories of, 157
Tilt, visual: aftereffect of, during lateral body tilt, 191–197
Transfer, bilingual: contextual vs. lexical meaning in, 28–32; familiarity with meaning and, 25 f., 28–31; phonological practice and, 31 f.; practice and, 23 ff., 28–39; scrambled sentences and, 28–31; speech errors in, delayed auditory feedback and, 32–35; spoonerisms in, 34 f.; word-order effects on, 26 f.

Understanding of implications, 367–373

Verticality, gravitational vs. visual: as reference for visual normalization, 191–197
Vision: aftereffects of tilt, lateral body tilt and, 191–197; normalization of, reference for, 191–197
Visual delay, luminance level and, 350–358
Visual feedback, changes in: motor-task performance and, 266 f., 269 ff.
Visual lightness: assimilation phenomena in, 56–72; contrast phenomena in, 56–72; Helson's theory of, 56 f., 69–72
Visual masking: cues in, 327–332; failure to replicate disinhibition of, 320–332
Visual normalization, retinal vs. gravitational references for, 191–197
Visual-orientation judgments, ocular countertorsion and, 195 ff.
Visual recognition, interaction between stored and sensory information in, 259–265
Visual search: type case and, 248–253; type size and, 248–253

Word identification: type case and, 248–253; type size and, 248–253
Word order, effect of: on bilingual transfer, 26 f.